TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

GEORGE DAVID BIRKHOFF

ARTHUR BYRON COBLE

EDWARD KASNER

WITH THE COÖPERATION OF

WILLIAM C. GRAUSTEIN	OLIVE C. HAZLETT	EINAR HILLE
WALLIE A. HURWITZ	DUNHAM JACKSON	AUBREY J. KEMPNER
HOWARD H. MITCHELL	CHARLES N. MOORE	ROBERT L. MOORE
FOREST R. MOULTON	CAROLINE E. SEELY	FRANCIS R. SHARPE
J. H. M. WEDDERBURN	ERNEST J. WILCZYNSKI	

VOLUME 26

1924

PUBLISHED BY THE SOCIETY

LANCASTER, PA., AND NEW YORK

1924

1

AN EXISTENCE THEOREM FOR THE CHARACTERISTIC NUMBERS OF A CERTAIN BOUNDARY VALUE PROBLEM*

BY

H. T. DAVIS

The first comprehensive study of the properties of the solutions of a linear, homogeneous, self adjoint differential equation of second order,

$$\frac{d}{dx}[k(x)\,u'(x)] + g(\lambda, x)\,u(x) = 0$$

and a set of boundary conditions was made by Sturm in 1836.† Sturm, limiting himself to a consideration of solutions which satisfied the boundary conditions

$$\alpha_1 u(a) + \alpha_2 u'(a) = 0, \qquad \beta_1 u(b) + \beta_2 u'(b) = 0,$$

established many fundamental theorems. One primary object was to find conditions upon the coefficients of the system sufficient to prove the existence of values of λ to which would correspond functions satisfying both differential equation and boundary conditions. These values of λ are called characteristic numbers. In his investigation Sturm was led to study the problem of how the zeros of these solutions depended upon the parameter and the coefficients of the system. These results have been called oscillation theorems.

Mason,‡ restricting $g(\lambda, x)$ to the form $\lambda A(x) - B(x)$, extended Sturm's existence theorem to the case of general linear boundary conditions with constant coefficients and gave an oscillation theorem for the special case

$$u(a) = u(b), \qquad u'(a) = u'(b).$$

Bôcher, employing other means, removed Mason's restrictions on $g(\lambda, x)$ for the oscillation problem, except that a certain uniqueness of the values of λ was not proved.§

* Presented to the Society, April 15, 1922.

† Journal de Mathématiques, vol. 1 (1836), pp. 106–186.

‡ These Transactions, vol. 7 (1906), pp. 337–360.

§ Paris Comptes Rendus, vol. 140 (1905), p. 928.

The next advance was made by Birkhoff who gave an existence proof and an oscillation theorem for the solution u_p of the self adjoint system of second order except that he limited himself to the case where $k = 1$ and the coefficients in his boundary conditions did not vary with λ. Birkhoff's method, like that of Bôcher, was based upon an application of Sturm's theorems to this system.*

Ettlinger completed the extension to the case where k varied with λ and the coefficients of the boundary conditions were functions of λ. The essential point of Ettlinger's treatment of the problem was the introduction of a condition which pertained to the sign of a certain determinant of fourth order made from the coefficients and their λ-increments.†

In a second paper,‡ Ettlinger used this condition in determining the oscillation properties of solutions belonging to a general set of linear boundary conditions.

The same problem with more or less generality has also been treated by Haupt§ and Fort,‖ the latter extending his results to a corresponding system in difference equations.

The general theory underlying the methods of Sturm and a complete formulation of the boundary value problem is to be found in Bôcher's *Leçons sur les Méthodes de Sturm* (Paris, 1917).

The object of the present paper is, first, to develop a general method of attack for the existence problem applicable to an extension of the Sturmian problem to equations of higher order than the second and, second, to apply this method to the second order problem from which follows a simplification of both argument and conditions.

The method will be found to be a slight modification of that employed by Bôcher, Birkhoff, and Ettlinger, use being made of the same principle of comparing one system with another by means of their characteristic equations.

I. The general problem

Suppose $L(u)$ is any linear, homogeneous, ordinary differential expression of nth order

$$L(u) = p_n(x)\frac{d^n u}{dx^n} + p_{n-1}(x)\frac{d^{n-1}u}{dx^{n-1}} + \cdots + p_0(x)\,u$$

* These Transactions, vol. 10 (1909), pp. 259–270.

† These Transactions, vol. 19 (1918), pp. 79–96.

‡ These Transactions, vol. 22 (1921), pp. 136–143.

§ Dissertation, Leipzig, Teubner, 1911. Also, Mathematische Annalen, vol. 76 (1914), pp. 67–104.

‖ American Journal of Mathematics, vol. 39 (1917), pp. 1–26.

and suppose that $p_n(x)$ is different from zero in the interval $a \leq x \leq b$. Consider, then, the two systems

$$\text{(I)} \quad \begin{aligned} L(u) &= 0, \\ U_i(u) &= 0, \end{aligned} \qquad \text{(II)} \quad \begin{aligned} L(u) &= 0, \\ V_i(u) &= 0 \end{aligned} \qquad (i = 1, 2, \ldots, n)$$

where U_i and V_i are expressions at two points a and b of the form

$$\alpha_{i1}u(a) + \alpha_{i2}u'(a) + \cdots + \alpha_{in}u^{(n-1)}(a) - \beta_{i1}u(b) - \beta_{i2}u'(b) - \cdots - \beta_{in}u^{(n-1)}(b).$$

Since the general solution of the differential equation may be written in the form

$$u = c_1 u_1 + c_2 u_2 + \cdots + c_n u_n,$$

where the functions $u_1, \ldots, u_n$ are a fundamental set of solutions of the equation, we are led to consider the following equations in order to determine the c_i for system I:

$$\begin{aligned} c_1 U_{11} + c_2 U_{12} + \cdots + c_n U_{1n} &= 0, \\ c_1 U_{21} + c_2 U_{22} + \cdots + c_n U_{2n} &= 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ c_1 U_{n1} + c_2 U_{n2} + \cdots + c_n U_{nn} &= 0, \end{aligned}$$

where the notation U_{ij} has been used to denote $U_i(u_j)$.

A similar set of conditions holds also for system II.

Consider, then, the two matrices

$$T_1 = \begin{Vmatrix} U_{11} \cdots U_{1n} \\ \cdots\cdots\cdots \\ U_{n1} \cdots U_{nn} \end{Vmatrix}, \qquad T_2 = \begin{Vmatrix} V_{11} \cdots V_{1n} \\ \cdots\cdots\cdots \\ V_{n1} \cdots V_{nn} \end{Vmatrix}.$$

The rank of a matrix is the maximum order of a determinant not zero issuing from the matrix. It is well known that if p is the rank of matrix T_1, there are $n - p$ linearly independent solutions for $c_1, \ldots, c_n$. Each of these solutions gives a function $u = c_1 u_1 + \cdots + c_n u_n$ which satisfies system I.*

* Bôcher, *Leçons*, Chap. II.

In what follows we shall suppose the ranks of both T_1 and T_2 to be $n-1$ and shall treat exceptions when they arise as special problems. We are led, thus, to consider the vanishing of two new expressions

$$F_1 = \begin{vmatrix} U_{11} \cdots U_{1n} \\ \cdots\cdots \\ U_{n1} \cdots U_{nn} \end{vmatrix}, \qquad F_2 = \begin{vmatrix} V_{11} \cdots V_{1n} \\ \cdots\cdots \\ V_{n1} \cdots V_{nn} \end{vmatrix}.$$

If the coefficients of $L(u)$, U_i, and V_i depend also upon a parameter λ, then F_1 and F_2 will be functions of λ, and for every value of λ for which F_1 or F_2 vanishes there will correspond a solution of system I or II. The equation $F_1(\lambda) = 0$ has been called by Bôcher the characteristic equation of system I.

We now assume that $F_1(\lambda)$ is known to vanish for a succession of values $\lambda_1, \lambda_2, \ldots . \lambda_n$. The object of this discussion is to develop conditions under which F_2 will vanish also.

Consider the new function

$$F(\lambda) = F_1 F_2, \qquad L_1 \leqq \lambda \leqq L_2.$$

If the coefficients of $L(u)$ are continuous throughout the region $a \leqq x \leqq b$, $L_1 \leqq \lambda \leqq L_2$, and the coefficients of U_i and V_i are continuous in the interval $(L_1 L_2)$, it may be shown that $F(\lambda)$ is also a continuous function in the interval $(L_1 L_2)$.* The following consequences are immediate:

Consider $F(\lambda)$ in the interval $\lambda_i \leqq \lambda \leqq \lambda_{i+1}$ where λ_i and λ_{i+1} are two successive zeros of F_1. If F_2 does not vanish in the interval, then the sign of $F(\lambda_i + \delta\lambda)$ will be the same as that of $F(\lambda_{i+1} - \delta\lambda)$. If, however, (a), the sign of $F(\lambda_i + \delta\lambda)$ and the sign of $F(\lambda_{i+1} - \delta\lambda)$ differ, then F_2 will have at least one zero between λ_i and λ_{i+1}. If, (b), for every such zero μ_i in the interval $(\lambda_i \lambda_{i+1})$ neither $F(\mu_i + \delta\lambda)$ nor $F(\mu_i - \delta\lambda)$ is zero, then the zeros μ_i must be isolated. Then, if μ_i and μ_{i+1} are successive zeros of F_2 in $(L_1 L_2)$ and if, (c), the signs of $F(\mu_i + \delta\lambda)$ and $F(\mu_{i+1} - \delta\lambda)$ differ, then F_1 will have at least one zero between the zeros of F_2. The condition (a), under the original assumption that F_1 is known to vanish at a succession of values $\lambda_1, \ldots, \lambda_n$, forms an existence theorem for system II and the conditions (a), (b), and (c) together assert that the characteristic numbers of systems I and II separate each other.

Now we have

$$F(\lambda_i + \delta\lambda) = (F_2 \cdot \delta F_1 + F_1 \cdot \delta F_2)_{\lambda = \lambda_i} + \text{terms of higher order}$$

* Bôcher, *Leçons*, pp. 7–8.

and since $F_1(\lambda_i) = 0$, we shall have to a first approximation

$$F(\lambda_i + \delta\lambda) = (F_2 \cdot \delta F_1)_{\lambda=\lambda_i} \text{ and } F(\lambda_{i+1} - \delta\lambda) = (F_2 \cdot \delta' F_1)_{\lambda=\lambda_{i+1}}.$$

Throughout this paper δ' will be used to denote the increment which results from a negative increment of the independent variable.

These results may now be stated in the following fundamental theorem:

Suppose $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ *are successive characteristic numbers of system* I *and* $(\mu_1, \mu_2, \ldots, \mu_n)$ *successive characteristic numbers of system* II. *If* $(F_2 \cdot \delta F_1)_{\lambda=\lambda_i}$ *and* $(F_2 \cdot \delta' F_1)_{\lambda=\lambda_{i+1}}$ *have different signs, then system* II *has at least one characteristic number between* λ_i *and* λ_{i+1}.

If $(F_1 \cdot \delta F_2)_{\lambda=\mu_i}$ *and* $(F_1 \cdot \delta' F_2)_{\lambda=\mu_{i+1}}$ *also have different signs, then the characteristic numbers of systems* I *and* II *alternate.*

II. Application to Sturm's system

The theorem developed in the last section will be applied first to the simple problem of finding conditions which will assure the existence of characteristic numbers of the following system of second order:

$$(1) \qquad \frac{d}{dx}[k(x)\,u'(x)] + g(\lambda, x)\,u(x) = 0, \qquad a \leqq x \leqq b,$$

$$(2) \qquad \begin{aligned} \alpha_1(\lambda)\,u(a) + \alpha_2(\lambda)\,u'(a) &= 0, \qquad -\infty \leqq \lambda \leqq +\infty, \\ \beta_1(\lambda)\,u(b) + \beta_2(\lambda)\,u'(b) &= 0. \end{aligned}$$

We shall assume $\alpha_i(\lambda)$ and $\beta_i(\lambda)$ to be continuous functions of λ, $k(x)$ to be continuous and positive in (ab) and $g(\lambda, x)$ a continuous, monotone increasing function in both arguments with the added condition that

$$(3) \qquad \lim_{\lambda=-\infty} g(\lambda, x) = -\infty, \qquad \lim_{\lambda=\infty} g(\lambda, x) = +\infty.$$

This condition may be shown to insure as large a number of zeros as we choose in the interval (ab) for any continuous solution $u(\lambda, x)$ by taking λ sufficiently large.* Moreover, due to the continuity conditions on the coefficients these zeros are continuous functions of λ except, perhaps, those at one extremity or the other of (ab).†

* Bôcher, *Leçons*, § 13.

† Bôcher, *Leçons*, p. 61.

Since there exists a solution of (1) which satisfies for all values of λ the initial conditions $u(a) = \alpha_2$, $u'(a) = -\alpha_1$, where $|\alpha_1| + |\alpha_2| \neq 0$, it is clear that for every value of λ there exists a solution which satisfies (1) and the first condition of (2). Assume, now, that α_1/α_2 is a non-decreasing function of λ. Clearly, then, α_2 could not vanish unless α_1 vanished also, but this case has already been excluded. Consequently, since $u(\lambda, x)$ will vanish in (ab) as many times as we please by choosing λ sufficiently large and since these zeros are continuous functions of λ, the zeros must enter the (ab) interval through b. Hence $u(\lambda, b)$ will vanish at an infinite number of values of λ, i. e. $(\lambda_1, \lambda_2, \ldots)$. In the special case $\alpha_2 \equiv 0$, no special assumption needs to be made since the zeros must necessarily enter through either $x = a$ or $x = b$.

Hence we see that for an infinite number of values solutions exist for the auxiliary system

$$\frac{d}{dx}[k(x)\, u'(x)] + g(\lambda, x)\, u(x) = 0,$$

$$\alpha_1 u(a) + \alpha_2 u'(a) = 0,$$

$$u(b) = 0.$$

In order to apply the theorem of the last section we must consider the characteristic functions $F_1(\lambda)$ and $F_2(\lambda)$, where the first belongs to the auxiliary system just written down and the second to system (1), (2).

It is clear that we may, without loss of generality, specialize a fundamental system of integrals $u_1(x)$, $u_2(x)$ by means of the relations

$$\begin{aligned} u_1(b) &= 0, \qquad \beta_1 u_1(b) + \beta_2 u_1'(b) = 1, \\ u_2(b) &= 1, \qquad \beta_1 u_2(b) + \beta_2 u_2'(b) = 0. \end{aligned} \tag{4}$$

In terms of these integrals the characteristic functions are easily seen to take the following simple form:

$$F_1 = \alpha_1 u_1(a) + \alpha_2 u_1'(a), \qquad F_2 = -[\alpha_1 u_2(a) + \alpha_2 u_2'(a)].$$

Since, by hypothesis,

$$F_1(\lambda_i) = 0, \tag{5}$$

we shall have

$$(\delta F_1\, F_2)_{\lambda=\lambda_i} = -(\alpha_1\, \delta u_1 + \alpha_2\, \delta u_1' + u_1\, \delta\alpha_1 + u_1'\, \delta\alpha_2)(\alpha_1\, u_2 + \alpha_2\, u_2')_{\lambda=\lambda_i}.$$

By means of the method of variation of parameters it may be shown that

$$\delta u_i^{(j)}(x) = P_i u_1^{(j)}(x) + Q_i u_2^{(j)}(x) - \frac{\beta_2}{k(b)} \int_b^x \delta g(t)\, u_i(t) \left[u_2(t)\, u_1^{(j)}(x) - u_1(t) u_2^{(j)}(x)\right] dt \qquad (i = 1, 2;\ j = 0, 1)$$

where we have set for brevity

$$P_i = -\beta_2 \left[\delta u_i(b)\, u_2'(b) - \delta u_i'(b)\, u_2(b)\right],$$

$$Q_i = -\beta_2 \left[\delta u_i'(b)\, u_1(b) - \delta u_i(b)\, u_1'(b)\right].$$

Consequently, making use of (5) and noting that $Q_1 = 0$, we have

$$(\delta F_1\ F_2)_{\lambda=\lambda_i} = \frac{\beta_2}{k(b)} \left[\int_a^b \delta g\, u_1^2\, dt\, (\alpha_1 u_2 + \alpha_2 u_2')^2 + \left(\frac{k(b)}{\beta_2}\right)^2 \left(\frac{\alpha_2^2}{k(a)}\right) \delta\left(\frac{\alpha_1}{\alpha_2}\right)\right].$$

Similarly

$$(\delta F_2\ F_1)_{\lambda=\mu_i} = -\frac{\beta_2}{k(b)} \left\{\left[-k(b)\, \delta\left(\frac{\beta_1}{\beta_2}\right) + \int_a^b \delta g\, u_2^2\, dt\right] (\alpha_1 u_1 + \alpha_2 u_1')^2 + \left(\frac{k(b)}{\beta_2}\right)^2 \left(\frac{\alpha_2^2}{k(a)}\right) \delta\left(\frac{\alpha_1}{\alpha_2}\right)\right\}.$$

From these two expressions, recalling the fundamental theorem, we may conclude that the characteristic numbers of the two systems will alternate if the following conditions are fulfilled:

(*a*) $g(\lambda, x)$ is a monotone increasing function of x and λ, satisfying condition (3);

(*b*) β_2/β_1 and α_1/α_2 are non-decreasing functions of λ;

(*c*) $|\beta_2|$ is bounded below;

(*d*) $|\alpha_1| + |\alpha_2| \neq 0$.

These conditions are the well known conditions for the existence of characteristic numbers for the Sturmian system (1), (2).

III. The general self adjoint system of second order

We shall consider next the general self adjoint system of second order,

$$\frac{d}{dx}[k(x)\,u'(x)] + g(\lambda, x)\,u(x) = 0,$$

$$\begin{aligned} \alpha_{11}(\lambda)\,u(a) + \alpha_{12}(\lambda)\,u'(a) &= \beta_{11}(\lambda)\,u(b) + \beta_{12}(\lambda)\,u'(b), \\ \alpha_{21}(\lambda)\,u(a) + \alpha_{22}(\lambda)\,u'(a) &= \beta_{21}(\lambda)\,u(b) + \beta_{22}(\lambda)\,u'(b), \\ k(a)\,(\beta_{11}\,\beta_{22} - \beta_{21}\,\beta_{12}) &= k(b)\,(\alpha_{11}\,\alpha_{22} - \alpha_{21}\,\alpha_{12}). \end{aligned} \tag{6}$$

We shall suppose that $k(x)$ and $g(\lambda, x)$ are limited by the same conditions as were imposed upon them in the last section. If now we let

$$L_{ij} = L_i(u_j) = \alpha_{i1}\,u_j(a) + \alpha_{i2}\,u_j'(a), \quad M_{ij} = M_i(u_j) = \beta_{i1}\,u_j(b) + \beta_{i2}\,u_j'(b),$$

where u_j is any particular solution of the differential equation, and choose for an auxiliary system the Sturmian system whose boundary conditions are

$$L_i(u) = 0, \qquad M_i(u) = 0,$$

then the functions F_1 and F_2 will be defined as follows:

$$F_1(\lambda) = \begin{vmatrix} L_{11} & L_{12} \\ M_{11} & M_{12} \end{vmatrix}, \qquad F_2(\lambda) = \begin{vmatrix} L_{11} - M_{11}, & L_{12} - M_{12} \\ L_{21} - M_{21}, & L_{22} - M_{22} \end{vmatrix}.$$

Let

$$\begin{aligned} L_{11} &= 0, \quad L_{21} = 1, \\ L_{12} &= 1, \quad L_{22} = 0, \\ \beta_{11}\,\beta_{22} - \beta_{21}\,\beta_{12} &= -k(b), \quad \alpha_{11}\,\alpha_{22} - \alpha_{21}\,\alpha_{12} = -k(a). \end{aligned} \tag{7}$$

It is always possible to do this since the first conditions merely particularize a fundamental set of solutions (u_1 and u_2), while the last condition amounts to multiplying the boundary conditions by a suitably chosen function of λ.

From these conditions, since $u_1 u_2' - u_2 u_1' = 1/k(x)$, it follows that

$$M_{11}\,M_{22} - M_{12}\,M_{21} = -1. \tag{8}$$

Consequently

$$\begin{aligned} F_1(\lambda) &= -M_{11}, \\ F_2(\lambda) &= M_{12} + M_{21} - 2, \end{aligned}$$

whence

$$(\delta F_1 F_2)_{\lambda=\lambda_i} = -\delta M_{11}(M_{12}+M_{21}-2)_{\lambda=\lambda_i}.$$

By means of the variation of parameters, it may be shown that

$$\delta u_i^{(j)}(x) = P_i u_1^{(j)}(x) + Q_i u_2^{(j)}(x) - \int_a^x \delta g(t)\, u_i(t)\,[u_1(t)\, u_2^{(j)}(x) - u_2(t)\, u_1^{(j)}(x)]\, dt \qquad (i=1,2;\ j=0,1),$$

where we have set for brevity

$$P_i = k(a)\,[\delta u_i(a)\, u_2'(a) - \delta u_i'(a)\, u_2(a)],$$

$$Q_i = k(a)\,[\delta u_i'(a)\, u_1(a) - \delta u_i(a)\, u_1'(a)].$$

Since $(M_{11})_{\lambda=\lambda_i} = 0$ and since, from the self adjoint condition on the coefficients

$$\begin{aligned} \alpha_{11}\,\delta\alpha_{22} + \alpha_{22}\,\delta\alpha_{11} - \alpha_{21}\,\delta\alpha_{12} - \alpha_{12}\,\delta\alpha_{21} &= 0, \\ \beta_{11}\,\delta\beta_{22} + \beta_{22}\,\delta\beta_{11} - \beta_{21}\,\delta\beta_{12} - \beta_{12}\,\delta\beta_{21} &= 0, \end{aligned} \tag{9}$$

we have

$$-(\delta F_1 F_2)_{\lambda=\lambda_i} = \left(Q_1 M_{12} - \int_a^b \delta g u_1^2\, dt\, M_{12}\right)(M_{12}+M_{21}-2) + \delta\beta_{11}\left(\frac{\beta_{12}}{k(b)} + \beta_{21}u_1^2 + \beta_{22}u_1u_1' - 2u_1\right) + \delta\beta_{12}\left(-\frac{\beta_{11}}{k(b)} + \beta_{21}u_1u_1' + \beta_{22}u_1'^2 - 2u_1'\right).$$

Let us recall from (8), since $M_{11}=0$, that $M_{12}M_{21}=1$. Hence

$$-(\delta F_1 F_2)_{\lambda=\lambda_i} = \left(Q_1 - \int_a^b \delta g u_1^2\, dt\right)(M_{12}-1)^2 + k(b)\,\delta\left(\frac{\beta_{11}}{\beta_{12}}\right)\left(u_1 - \frac{\beta_{12}}{k(b)}\right)^2. \tag{10}$$

Similarly

$$-(\delta F_2 F_1)_{\lambda=\mu_i} = (\delta M_{12} + \delta M_{21})\, M_{11}.$$

Since, from (7) and (9), we get $P_1 + Q_2 = 0$, and since

$$F_2(\mu_i) = (M_{12}+M_{21}-2)_{\lambda=\mu_i} = 0, \tag{11}$$

it follows that

$$-(\delta F_2 F_1)_{\lambda=\mu_i} = \left(P_2 - \int_{22}\right) M_{11}^2 + 2\left(Q_2 + \int_{21}\right) M_{11}(M_{12}-1)$$

$$-\left(Q_1 + \int_{11}\right)(1-M_{12})^2 + (u_2\delta\beta_{11} + u_2'\delta\beta_{12} + u_1\delta\beta_{21} + u_1'\delta\beta_{22})\, M_{11},$$

where we have used the abbreviation $\int_{ij} = \int_a^b -\delta g u_i u_j\, dt$.

But by the use of (8) and (11) and the fact that $u_1u_2' - u_2u_1' = 1/k(x)$, it may be shown that

$$(u_2\delta\beta_{11} + u_2'\delta\beta_{12} + u_2\delta\beta_{21} + u_1'\delta\beta_{22})\, M_{11} = (\beta_{11}\delta\beta_{21}$$

$$-\beta_{21}\delta\beta_{11})\, u_1^2 + 2(\beta_{12}\delta\beta_{21} - \beta_{22}\delta\beta_{11})\, u_1u_1' + (\beta_{12}\delta\beta_{22}$$

$$-\beta_{22}\delta\beta_{12})\, u_1'^2 + 2\delta\beta_{11}u_1 + 2\delta\beta_{12}u_1' + (\beta_{11}\delta\beta_{12} - \beta_{12}\delta\beta_{11})/k(b).$$

In order to show that this quadratic polynomial is always positive, it will be sufficient to know that D_3, the determinant of the coefficients, is positive or zero and that D_2 and D_1, the leading principal minors of the determinant, are both positive. Similarly the polynomial will be negative if D_3 is negative or zero, D_2 positive, and D_1 negative.

By use of the second equation in (9) and by a proper combination of multiples of columns in D_3, it follows readily that

$$D_1 = \beta_{11}\delta\beta_{21} - \beta_{21}\delta\beta_{11}, \quad D_2 = k(b)(\delta\beta_{12}\delta\beta_{21} - \delta\beta_{22}\delta\beta_{11}), \quad D_3 = 0.$$

Similarly the determinant of the quadratic form in M_{11} and $(M_{12}-1)$ is

$$D_4 = \left(Q_2 + \int_{21}\right)^2 + \left(P_2 - \int_{11}\right)\left(Q_1 + \int_{11}\right)$$

$$= Q_2^2 + P_2Q_1 + \int_a^b -\delta g(-Q_1u_2^2 + 2Q_2u_1u_2 + P_2u_1^2)\, dt + \int_{21}^2 - \int_{11}\int_{22}.$$

From these determinants and from (10), it is evident that the theorem of the first section will apply providing the following conditions are imposed, in which, as before, the prime pertains to the increment resulting from a negative increment of the independent variable:

At

$$\lambda = \lambda_i; \qquad Q_1 < 0, \qquad \delta\left(\frac{\beta_{12}}{\beta_{11}}\right) \geqq 0,$$

$$\lambda = \lambda_{i+1}; \qquad Q_1' > 0, \qquad \delta'\left(\frac{\beta_{12}}{\beta_{11}}\right) \leqq 0,$$

$$(12)\quad \lambda = \mu_i; \quad Q_2^2 + P_2 Q_1 \leqq 0, \quad Q_1 < 0,$$

$$\delta\left(\frac{\beta_{21}}{\beta_{11}}\right) \geqq 0, \quad \delta\beta_{11}\,\delta\beta_{22} - \delta\beta_{21}\,\delta\beta_{12} \leqq 0,$$

$$\lambda = \mu_{i+1};\ Q_2'^2 + P_2' Q_1' \leqq 0, \quad Q_1' > 0,$$

$$\delta'\left(\frac{\beta_{21}}{\beta_{11}}\right) \leqq 0, \quad \delta'\beta_{11}\,\delta'\beta_{22} - \delta'\beta_{21}\,\delta'\beta_{12} \leqq 0.$$

Since $Q_2^2 + P_2 Q_1 = (1/k(a))\,(\delta\alpha_{22}\,\delta\alpha_{11} - \delta\alpha_{12}\,\delta\alpha_{21})$ and $Q_1 = (\alpha_{11}^2/k(a)) \cdot \delta\,(\alpha_{12}/\alpha_{11})$, we may state the following existence theorems:

THEOREM I. *Between two successive characteristic numbers of the Sturmian system $L_1 = 0$, $M_1 = 0$, there must always lie at least one characteristic number for the general self adjoint system of second order $L_1 = M_1$, $L_2 = M_2$.*

THEOREM II. *Under the additional conditions*

(1) $\dfrac{\beta_{21}}{\beta_{11}}$ *a non-decreasing function,* $|\beta_{11}| + |\beta_{21}| \neq 0$;

(2) $\delta\beta_{11}\,\delta\beta_{22} - \delta\beta_{21}\,\delta\beta_{12}$ *and* $\delta'\beta_{11}\,\delta'\beta_{22} - \delta'\beta_{21}\,\delta'\beta_{12} \leqq 0$,

$\delta\alpha_{11}\,\delta\alpha_{22} - \delta\alpha_{21}\,\delta\alpha_{12}$ *and* $\delta'\alpha_{11}\,\delta'\alpha_{22} - \delta'\alpha_{21}\,\delta'\alpha_{12} \leqq 0$,

the characteristic numbers of the Sturmian system and the general self adjoint system alternate.

It must be noted that both of these theorems are valid only under the assumption that the rank of the matrix of the general system is different from zero. This exception is the problem treated in the next section.

IV. The case of double values

We observe, first, that when $M_{11} = 0$, $M_{22} = 0$, $M_{12} = 1$, $M_{21} = 1$, the rank of the matrix T_2 (Section I) is zero and consequently two linearly independent solutions must exist for such a value of λ. Let μ be such a double value of λ.

Then

$$(\delta F)_{\lambda=\mu} = 0, \quad (\delta^2 F)_{\lambda=\mu} = 0, \quad (\delta^3 F)_{\lambda=\mu} = 3\,(\delta F_1\, \delta^2 F_2)_{\lambda=\mu},$$

$$(\delta F_1)_{\lambda=\mu} = -(\delta M_{11})_{\lambda=\mu} = -Q_1 + \int_a^b \delta g\, u_1^2\, dt + \frac{\beta_{12}^2}{k(b)}\, \delta\left(\frac{\beta_{12}}{\beta_{11}}\right).$$

Consequently, under the conditions already imposed, F_1 changes sign at μ. We next compute

$$(\delta^2 F_2)_{\lambda=\mu} = 2\,(\delta M_{22}\, \delta M_{11} - \delta M_{12}\, \delta M_{21})_{\lambda=\mu}$$

$$= 2\left(\sum_{i=1}^{6} A_i\, U_i\right)$$

where the A_i are the determinants in order that can be formed from the matrix

$$\left\| \begin{matrix} \beta_{11} & \beta_{12} & \delta\beta_{11} & \delta\beta_{12} \\ \beta_{21} & \beta_{22} & \delta\beta_{21} & \delta\beta_{22} \end{matrix} \right\|$$

and the U_i are the corresponding determinants from the matrix

$$\left\| \begin{matrix} \delta u_1 & \delta u_1' & u_1 & u_1' \\ \delta u_2 & \delta u_2' & u_2 & u_2' \end{matrix} \right\|.$$

From a computation based on the formulas of the last section it may be proved that

$$U_1 = -\frac{1}{k(b)}\, D_4\,; \qquad U_6 = \frac{1}{k(b)}\,.$$

Upon expansion $A_2 U_2 + A_3 U_3 + A_4 U_4 + A_5 U_5$ is found to be a quadratic form in u_1, u_2, u_1', u_2' whose matrix is the following symmetric array:

$$\begin{Vmatrix} A_2 K & A_3 K & A_2 M & A_3 M \\ A_3 K & A_5 K & A_3 M & A_5 M \\ A_2 M & A_3 M & A_2 N & A_3 N \\ A_3 M & A_5 M & A_3 N & A_5 N \end{Vmatrix}$$

where

$$K = -P_2 - \int_a^b \delta g u_2^2 \, dt, \quad M = -Q_2 + \int_a^b \delta g u_1 u_2 \, dt, \quad N = Q_1 - \int_a^b \delta g u_1^2 \, dt.$$

Suppose Δ_4 is the determinant of the matrix and Δ_1, Δ_2, Δ_3 are the leading principal minors of Δ_4. Then the quadratic form will be positive definite providing $\Delta_1, \ldots, \Delta_4$ are all positive, or negative definite if $\Delta_1, \ldots, \Delta_4$ are alternately negative and positive.

It can be proved upon expansion that $\Delta_1, \ldots, \Delta_4$ are resolvable into the following factors:

$$\Delta_1 = A_2 K, \qquad \Delta_3 = A_2 K (A_2 A_5 - A_3^2)(KN - M^2),$$

$$\Delta_2 = K^2 (A_2 A_5 - A_3^2), \quad \Delta_4 = (A_2 A_5 - A_3^2)^2 (KN - M^2)^2.$$

Consequently we must demand that

$$A_2 A_5 - A_3^2 \geqq 0, \qquad KN - M^2 \geqq 0$$

and the sign of the form will be that of $A_2 K$.

But

$$KN - M^2 = -D_4 \text{ (of the preceding section)},$$

$$A_2 A_5 - A_3^2 = D_2,$$

$$A_2 = D_1.$$

Therefore, since it follows from (12) that $A_2 K \leqq 0$, the quadratic form must be negative definite under the restrictions imposed by Theorem II, from which it follows that $\delta^2 F_2 < 0$.

THEOREM III. *Under the conditions of Theorem* II, *two successive zeros of* F_2 *are always separated by a zero of* F_1. *As a limiting case two zeros of* F_2 *may coincide with a zero of* F_1. *Since the sign of* $F = F_1 F_2$ *changes at every zero of* F_1 *and* F_2, *the ordering of the zeros is uniquely determined from a knowledge of the sign of* F *when* λ *approaches negative infinity.*

V. EXTENSION TO THE CASE WHERE k VARIES WITH λ

The conclusions just obtained afford an easy method of extension to the case where k is a function of λ as well as of x.

If we transform the general system (6) by means of the transformation $u = Q(x)v$ where $Q(x) = 1/\sqrt{k(x)}$, we shall obtain the following system

$$v'' + q(\lambda, x)\, v = 0,$$

$$A_{11}v(a) + A_{12}v'(a) = B_{11}v(b) + B_{12}v'(b), \tag{13}$$

$$A_{21}v(a) + A_{22}v'(a) = B_{21}v(b) + B_{22}v'(b),$$

where $q(\lambda, x) = Q[(kQ')' + gQ]$, and

$$A_{i1} = \alpha_{i1}Q(a) + \alpha_{i2}Q'(a), \qquad A_{i2} = \alpha_{i2}Q(a),$$

$$B_{i1} = \beta_{i1}Q(b) + \beta_{i2}Q'(b), \qquad B_{i2} = \beta_{i2}Q(b) \qquad (i = 1, 2).$$

A simple computation will show that the new system is also self adjoint.

Now the values of λ which furnish a solution of this system are evidently values which furnish a solution of system (6) since we may pass from one to the other by means of a non-singular transformation.

But system (13) is itself included under the theorems which apply to system (6) so that if the coefficients of (13) are properly restricted the existence of λ's for this system and consequently for system (6) will be established. Since the k is now absorbed in $q(\lambda, x)$ we need no longer restrict it to be a function independent of λ.

We should notice also that, since Q does not vanish in the interval, the oscillation properties of u and v are identical.

VI. THE ORDERING OF THE ZEROS OF F_1 AND F_2

We saw from Theorem III that the ordering of the zeros of F_1 and F_2 is uniquely determined from a knowledge of the sign of F when λ approaches

negative infinity. We consider $F = F_1 F_2 = -M_{11} M_{12} - M_{11} M_{21} + 2 M_{11}$ as $\lambda \to -\infty$.

From the relation $u_1 u_2' - u_2 u_1' = 1/k(x)$ and recalling (7) we may prove that

$$u_2(b) = \left[-\frac{\alpha_{22}}{\alpha_{12}} + \int_a^b \frac{dt}{k(t)\, u_1^2(t)} \right] u_1(b) \qquad = S\, u_1(b),$$

$$u_2'(b) = \left[-\frac{\alpha_{22}}{\alpha_{12}} + \int_a^b \frac{dt}{k(t)\, u_1^2(t)} \right] u_1'(b) + \frac{1}{k(b)\, u_1(b)} = S\, u_1'(b) + T.$$

Under the conditions already imposed upon u_1 it may be shown* that for $x > a$,

$$\lim_{\lambda=-\infty} \frac{u_1'(x)}{u_1(x)} = \infty, \quad \lim_{\lambda=-\infty} u_1'(x) = \infty, \quad \lim_{\lambda=-\infty} u_1(x) = \infty.$$

Substituting the values for $u_2(b)$ and $u_2'(b)$ in F we have

$$F = -[(\beta_{11}^2 S + \beta_{11}\beta_{21})\, u_1^2 + (\beta_{11}\beta_{22} + \beta_{12}\beta_{21} + 2\beta_{11}\beta_{12} S)\, u_1 u_1'$$
$$+ (\beta_{12}^2 S + \beta_{12}\beta_{22})\, u_1'^2 + (\beta_{11}\beta_{12} T - 2\beta_{11})\, u_1 + (\beta_{12}^2 T - 2\beta_{12})\, u_1'].$$

If we notice that the discriminant of the first three terms of F, since $\beta_{21}\beta_{12} - \beta_{11}\beta_{22} = k(b)$, is equal to $k^2(b)$, it may be verified that F is factorable as follows:

$$F = -\pi[B\pi + k(b)] \left[u_1 - \frac{\beta_{12} + \beta_{12}\sqrt{\dfrac{-B\pi}{k(b)}}}{B\pi + k(b)} \right] \left[u_1 - \frac{\beta_{12} + \beta_{12}\sqrt{\dfrac{-B\pi}{k(b)}}}{B\pi + k(b)} \right],$$

where $B = \beta_{12}^2 S + \beta_{12}\beta_{21}$, and $\pi = u_1'/u_1 + \beta_{11}/\beta_{12}$.

If we make the restriction that β_{11} and β_{12} shall be bounded when $\lambda \to -\infty$, recalling from Section II that β_{12} can never vanish, we see that the sign of F depends entirely upon the sign of B, being positive when $B < 0$, and negative when $B \geqq 0$.

* Bôcher, *Leçons*, pp. 65–66.

It is important to notice that we have here assumed that $\alpha_{12} \neq 0$, but if this were not the case then we could express $u_1(b)$ in terms of $u_2(b)$, choose for our comparison system $F = M_{22}$, and obtain a similar result.

These results, recalling Theorems II and III, may be stated as follows:

THEOREM IV. *If β_{11} and β_{12} are bounded as λ approaches negative infinity, then the first characteristic number of the Sturmian system precedes the first characteristic number of the general self adjoint system provided* $\lim_{\lambda=-\infty} B \geqq 0$; *but the first characteristic number of the general system precedes or (in the case of double values) coincides with the first characteristic number of the Sturmian system provided* $\lim_{\lambda=-\infty} B < 0$.

INDIANA UNIVERSITY,
BLOOMINGTON, IND.

A THEOREM ON THE FACTORIZATION OF POLYNOMIALS OF A CERTAIN TYPE*

BY

LLOYD L. DINES

Polynomials of the form

$$P \equiv y^n + P_{n-1}(x)\, y^{n-1} + \cdots + P_1(x)\, y + P_0(x),$$

where the coefficients $P_j(x)$ are power series converging in a sufficiently restricted neighborhood of the origin and vanishing with x, are of importance in the theory of implicit functions. The form of analytic expression of the functions $y(x)$ defined by the equation $P = 0$ depends upon the *characteristic line* (often called the Newton polygon) of the polynomial P. The nature of this dependence strongly suggests the theorem of the present paper, and may in fact afford a proof of it. The proof here given, however, proceeds along more direct lines, and like the theorem itself is algebraic in character. An analogous theorem, relative to polynomials of which the coefficients are "Hensel series" proceeding according to powers of a prime number p, and for which the factorization is to be made in this domain was proved by Dumas.† The theorem of the present paper, like that of Dumas, admits a converse, and both converses are special cases of the "Theorem on the Product" proved by Blumberg‡ for a very general class of polynomials defined postulationally.

1. THE THEOREM

If the characteristic line§ of the polynomial P is a broken line of k segments, then P is a product of k polynomials of the same type as P, each of which has

* Presented to the Society, September 7, 1923.

† Journal de Mathématiques, ser. 6, vol. 2 (1906), p. 223.

‡ These Transactions, vol. 17 (1916), p. 530.

§ The characteristic line is to be thought of as drawn relative to a system of rectangular i, j axes, its form being determined by the "degree points" (i, j) for which the coefficients p_{ij} of the double series

$$P \equiv \sum_{i=0}^{\infty} \sum_{j=0}^{n} p_{ij}\, x^i\, y^j$$

are different from zero. It is not, strictly speaking, an entire "Newton polygon", but only that part which is convex toward the origin $i = 0$, $j = 0$. For a definition and discussion of the Newton polygon, see Chrystal's *Algebra*, Part II, pp. 386–96.

an unbroken characteristic line the same in length and direction as one segment of the characteristic line of P.

This theorem is evidently an immediate consequence, by repetition, of the following somewhat simpler proposition:

If the characteristic line of P is a broken line, and if m is the ordinate of its lowest vertex, then P is a product of two polynomials of the same type as P, of degrees m and $n-m$ respectively, and such that their characteristic lines can be obtained by translation from the two sections of the characteristic line of P of which this vertex is the upper end and the lower end respectively.

We will prove the proposition in this simpler form. To do this we must prove the existence of two polynomials

$$A \equiv y^m + A_{m-1}(x)\, y^{m-1} + \cdots + A_1(x)\, y + A_0(x),$$

$$B \equiv y^\mu + B_{\mu-1}(x)\, y^{\mu-1} + \cdots + B_1(x)\, y + B_0(x) \quad (\mu = n-m),$$

satisfying the identity

$$A \cdot B \equiv P. \tag{1}$$

Substituting for A, B, and P in (1) their polynomial expressions, and equating coefficients of like powers of y on the two sides of the equality, we obtain a system of equations which may be written

$$A_j B_0 + A_{j-1} B_1 + \cdots + A_{j-r} B_r + \cdots + A_0 B_j = P_j(x) \qquad (j = 0, 1, \ldots, n-1), \tag{2}$$

with the following interpretation of notations:

$$A_m = B_\mu = 1, \qquad A_{m+q} = B_{\mu+q} = 0 \qquad (q > 0). \tag{3}$$

The system (2) consists of n equations for the determination of the n coefficients

$$A_j \quad (j = 0, 1, \ldots, m-1), \qquad B_j \quad (j = 0, 1, \ldots, \mu-1).$$

We will show that under the conditions of the hypothesis upon the characteristic line of P, the determination is possible, and unique.

2. A TRANSFORMATION OF THE SYSTEM (2)

We denote by the symbol $[j]$, the smallest integer i for which the point (i, j) does not lie to the left of the characteristic line of P. As a consequence of this definition, each coefficient $P_j(x)$ contains $x^{[j]}$ as a factor. We introduce the quantities $\bar{P}_j$, $\bar{A}_j$, and $\bar{B}_j$ as follows:

$$(4)\quad P_j \equiv x^{[j]}\bar{P}_j,\quad A_j \equiv x^{[j]-[m]}\bar{A}_j,\quad \bar{B}_j \equiv x^{[j+m]}\bar{B}_j \quad (j = 0, 1, \ldots, n-1).$$

Substituting these values of P_j, A_j, B_j into the system of equations (2), and dividing the jth equation by $x^{[j]}$ for each j, we obtain a system which may be written

$$(5)\quad \bar{A}_j\bar{B}_0 + x^{\{j,1\}}\bar{A}_{j-1}\bar{B}_1 + \cdots + x^{\{j,r\}}\bar{A}_{j-r}\bar{B}_r + \cdots + x^{\{j,j\}}\bar{A}_0\bar{B}_j = \bar{P}_j(x) \quad (j = 0, 1, \ldots, n-1),$$

where the symbol $\{j, r\}$ has the meaning

$$\{j, r\} \equiv [j-r]-[m]+[m+r]-[j].$$

We shall see that the exponent $\{j, r\}$ is never negative in a non-vanishing term of (5), but that in a certain well defined class of terms it is indeed positive.

3. THE EXPONENT $\{j, r\}$ AND THE ARTIFICIAL ORDER

Let us denote by k_α and m_α, respectively, the projections upon the i and j axes of that segment of the characteristic line of P which meets and lies above the line $j = \alpha$. For the sake of brevity, we will denote the projections of the lowest segment merely by k and m. That is, when $\alpha < m$, $k_\alpha = k$, and $m_\alpha = m$. And from the defining properties of the characteristic line it follows that if $\alpha \geqq m$, then $k_\alpha/m_\alpha < k/m$.

From the definition of the symbol $[j]$ and the properties of the characteristic line follow certain inequalities which we shall need, and which may be stated as follows:

$$(6)\quad [m-q]-[m] \geqq q\frac{k}{m}, \qquad 0 \leqq q \leqq m,$$

$$(7)\quad [m]-[m+r] < r\frac{k}{m}, \qquad 0 < r \leqq \mu,$$

$$(8)\quad [j]-[j+s] < s\frac{k_j}{m_j} + 1, \qquad 0 \leqq s \leqq m+\mu-j.$$

We are now in a position to study the exponent $\{j, r\}$. Since by (3) and (4) $\bar{A}_{j-r} = 0$ when $j-r > m$, this exponent can occur in non-vanishing terms of (5) only if $j-r \leq m$, that is if $j \leq m+r$.

Taking $q = m-(j-r)$ in (6), and $s = (m+r)-j$ in (8) we obtain the two relations

$$[j-r]-[m] \geqq (m-j+r)\frac{k}{m}, \tag{6'}$$

$$[j]-[m+r] < (m+r-j)\frac{k_j}{m_j}+1. \tag{8'}$$

Subtracting the second of these relations from the first, we find that

$$\{j, r\} > (m-j+r)\left(\frac{k}{m}-\frac{k_j}{m_j}\right)-1,$$

and since the right hand side of this inequality cannot be less than -1, we have shown that *the exponent* $\{j, r\}$ *cannot be negative in a non-vanishing term of* (5).

Certain other important statements to be made relative to the exponent $\{j, r\}$ depend upon what we shall call the "artificial order" for the subscripts $j = 0, 1, 2, \ldots, n-1$. This artificial order is suggested by the rôle which the lowest segment of the characteristic line plays in our theorem. If the coördinates of the upper and lower ends of this last segment be denoted by (l, m) and $(k+l, 0)$ respectively, the equation of the last segment is

$$mi+kj = (k+l)m.$$

By the artificial order of the subscripts j, we shall mean the order in which the corresponding points $([j], j)$ occur on the sequence of parallel lines

$$mi+kj = \alpha, \qquad \alpha = 1, 2, 3, \ldots.$$

the order on each individual line being downward.

More precisely, we define the *level* of a subscript j to be the integer given by the formula

$$\textit{level of } j = m[j]+kj.$$

We then order the subscripts $j = 0, 1, \ldots, n-1$, according to increasing level, with the additional convention that of two subscripts having the same level, the greater subscript shall precede.

In the statements (a) and (b) which follow, the term "order" is to be interpreted as the artificial order just defined.

(a) If the order of j is lower than the order of $j-r$, then $\{j, r\} > 0$.

(b) If $j < m+r$, and the order of j is lower than the order of $m+r$, then $\{j, r\} > 0$.

To prove (a), we have by hypothesis

$$m\,[j] + kj \leqq m\,[j-r] + k\,(j-r)$$

and hence

$$[j-r] - [j] \geqq r\,\frac{k}{m}.$$

This inequality, together with (7), yields the conclusion of (a), namely $\{j, r\} > 0$.

To prove (b), we have by hypothesis

$$m\,[j] + kj < m\,[m+r] + k\,(m+r)$$

or

$$[m+r] - [j] > (j-m-r)\,\frac{k}{m}.$$

From this inequality, and the inequality (6′) follows the conclusion of (b).

4. The solution of the system of equations

Turning now to the system (5), we note that since $\bar{A}_j = 0$ when $j > m$, the equations fall under two slightly different types, namely

$$\bar{A}_j B_0 + \sum_{r=1}^{j} x^{\{j,r\}}\,\bar{A}_{j-r}\,\bar{B}_r = \bar{P}_j(x) \qquad (j < m),$$

$$\bar{B}_{j-m} + \sum_{r=j-m+1}^{j} x^{\{j,r\}}\,\bar{A}_{j-r}\,\bar{B}_r = \bar{P}_j(x) \qquad (j \geqq m). \tag{9}$$

We will hereafter think of the equations of this system as arranged in the order indicated by the artificial order of the subscripts j occurring on the right hand sides. Furthermore the n unknowns $\bar{A}_j$ ($j = 0, 1, \ldots, m-1$)

and $\bar{B}_r$ $(r = 0, 1, \ldots, \mu - 1)$ are to be ordered according to the following conventions: $\bar{A}_j$ is to have the same place in the sequence of unknowns that its subscript j has in the artificial order of subscripts; $\bar{B}_r$ is to have the same place in the sequence of unknowns that $m + r$ has in the artificial order of subscripts.

By virtue of this ordering of unknowns, it follows from (a) and (b) that the equations (9) can be written in the forms

$$\begin{aligned} G_j + \bar{A}_j \bar{B}_0 + x H_j &= \bar{P}_j(x) && (j < m), \\ G_j + \bar{B}_{j-m} + x H_j &= \bar{P}_j(x) && (j \geq m), \end{aligned} \tag{10}$$

where G_j and H_j are binary forms in the unknowns A and B, and furthermore each G_j has the important property that it contains no A or B of order as high as the order of the subscript j.

Consider now the system of equations obtained from (10) by putting $x = 0$:

$$\begin{aligned} G_j + \bar{A}_j \bar{B}_0 &= \bar{P}_j(0) && (j < m), \\ G_j + \bar{B}_{j-m} &= \bar{P}_j(0) && (j \geq m). \end{aligned} \tag{11}$$

Since the first subscript j in the artificial order is $j = m$,* the first equation of this system is

$$\bar{B}_0 = \bar{P}_m(0).$$

The value of $\bar{B}_0$ so determined is different from zero, since $\bar{P}_m(0) = p_{lm} \neq 0$.

Since, as we have seen, each form G_j contains only $\bar{A}$'s and $\bar{B}$'s of order lower than the order of the corresponding subscript j, and since $\bar{B}_0 \neq 0$, it follows that the equations (11) can be solved successively in their artificial order for $\bar{A}_j$ $(j = 0, 1, \ldots, m-1)$ and $\bar{B}_{j-m}$ $(j = m, m+1, \ldots, n-1)$. We thus obtain a solution of the system (10) corresponding to the value $x = 0$, which we may denote by

$$x = 0, \qquad \bar{A}_j = a_j, \qquad \bar{B}_{j-m} = b_{j-m}. \tag{12}$$

The fundamental existence theorem for implicit functions insures the existence of power series $\bar{A}_j(x)$, $B_{j-m}(x)$ converging in a neighborhood of

* The point (l, m) or $([m], m)$ being the upper end of the lowest segment of the characteristic line.

the origin, and identically satisfying the system of equations (10), provided the functional determinant of the system with respect to the n variables $\bar{A}_j$, $\bar{B}_{j-m}$ is different from zero for the set of values (12). We therefore consider now this functional determinant.

If the order of the rows of this determinant be made to agree with the artificial order of the subscripts j, and the order of the columns to agree with the artificial order of the variables $\bar{A}$ and $\bar{B}$, it is apparent that the elements of the principal diagonal will for $x = 0$ be $b_0\,(= \bar{B}_0)$ or 1, according as the corresponding subscript j is less than m or not. Furthermore every element to the right of the principal diagonal is zero for $x = 0$. Hence the functional determinant has for the set of values (12) the value

$$\Delta = b_0^m = p_{lm}^m \neq 0.$$

The system of equations (10) have therefore by the existence theorem, unique convergent solutions

$$\bar{A}_j = a_j + \mathfrak{A}_j(x) \qquad (j = 0, 1, \ldots, m-1), \tag{13}$$

$$\bar{B}_{j-m} = b_{j-m} + \mathfrak{B}_{j-m}(x) \quad (j = m, m+1, \ldots, n-1). \tag{14}$$

In view of the substitutions (4), the system (2) has therefore the unique solution

$$A_j = x^{[j]-l}(a_j + \mathfrak{A}_j(x)) \qquad (j = 0, 1, \ldots, m-1), \tag{15}$$

$$B_j = x^{[j+m]}(b_j + \mathfrak{B}_j(x)) \qquad (j = 0, 1, \ldots, \mu-1). \tag{16}$$

5. The characteristic lines of A and B

We have shown the existence of factors

$$A \equiv y^m + A_{m-1}(x)\, y^{m-1} + \cdots + A_0(x),$$

$$B \equiv y^\mu + B_{\mu-1}(x)\, y^{\mu-1} + \cdots + B_0(x)$$

satisfying the identity (1). It remains to show that the characteristic lines of A and B can be obtained from parts of the characteristic line of P by suitable translation.

We first recall that the characteristic line of P passes through or to the left of all points $([j], j)$ for $j = 0, 1, \ldots, n-1$.

If the lowest segment of this line be translated l units to the left, it will pass through or to the left of all points $([j]-l, j)$ for $j \leq m$. From the analytic form of the coefficients (15) it follows therefore that the characteristic line of A will certainly not lie to the left of this translated segment. But it is easy to see that the ends (l, m) and $(k+l, 0)$ of the segment subjected to translation go into points of the characteristic line of A. The former point goes into $(0, m)$ which is obviously the upper end of the characteristic line. The latter goes into $(k, 0)$ which will be the lower end of this characteristic line provided $a_0 \neq 0$. This is true since a_0 is determined by the equation $a_0 b_0 = \overline{P}_0(0) = p_{k+l,0}$ and $p_{k+l,0} \neq 0$. The desired property relative to the characteristic line of A is thus established.

Consider now the part of the characteristic line of P above the point (l, m). If it be translated m units downward it will pass through or to the left of all points $([j+m], j)$ for $j \leq \mu$. Hence from (16), no part of the characteristic line of B will lie to the left of its new position. Furthermore the characteristic line of B must coincide with it in its new position provided in (16) $b_j \neq 0$ for every j for which $([j], j)$ is a vertex of the characteristic line of P.

Suppose then that $([j], j)$ is a vertex for P. In this case, the inequality (8′) may be altered to the form

$$[j] - [m+r] \leq (m+r-j)\frac{k_j}{m_j},$$

and as a result $\{j, r\} > 0$, at least for $j < m+r$. Hence when $([j], j)$ is a vertex, the form G_j in (11) contains no terms, and the solution of the corresponding equation is $b_{j-m} = \overline{P}_j(0) = p_{[j],j} \neq 0$. Hence the b_{j-m} of (14), that is the b_j of (16), is different from zero when $([j], j)$ is a vertex for P. This completes the proof of the theorem.

UNIVERSITY OF SASKATCHEWAN,
SASKATOON, CANADA.

A FUNDAMENTAL CLASS OF GEODESICS ON ANY CLOSED SURFACE OF GENUS GREATER THAN ONE*

BY

HAROLD MARSTON MORSE

INTRODUCTION

1. The study of geodesics on closed surfaces is influenced largely by the genus of the surface. Of closed surfaces of genus zero an important class has been studied by H. Poincaré,† namely, those closed surfaces which are everywhere convex. Of the surfaces with genus one, the torus has been studied by G. A. Bliss,‡ and it seems probable that the types of geodesics found there will be found among the geodesics on any closed surfaces of genus one. It is the object of this paper to consider geodesics on any closed surface of genus greater than one.

In studying geodesics on a surface of negative curvature, the author§ found that, of those geodesics which if extended indefinitely in either sense remained in a finite part of space, any particular one could be characterized in a manner which depended only upon a succession of fundamental contours of the surface. Now these surfaces of negative curvature are never closed. The question accordingly arose, is it possible also on closed surfaces, to characterize the geodesics in terms of the topographical elements of the surface, that is to characterize the geodesics of any particular closed surface in terms which would serve likewise for any other closed surface of the same genus? The answer to this question was readily seen to be no. However, on surfaces of genus $p > 1$, there appeared a fundamental class of geodesics which could be identified once and for all for all closed surfaces of the same genus. This paper is devoted to the definition and study of such a fundamental class of geodesics.

The surface is first mapped upon a hyperbolic non-euclidean plane. In the theory as developed each non-euclidean straight line represents a different

* Presented to the Society, September 8, 1921.

† *Sur les lignes géodésiques des surfaces convexes,* these Transactions, vol. 6 (1905), p. 237.

‡ *The geodesic lines on the anchor ring,* Annals of Mathematics, ser. 2, vol. 4 (1903), p. 1.

§ *Recurrent geodesics on a surface of negative curvature,* these Transactions, vol. 22 (1921), p. 84.

type of fundamental geodesic and it is with the aid of this representation that the fundamental class of geodesics is studied.

THE SURFACE

2. **The surface defined.** The surfaces to be considered are to be two-sided, closed surfaces of finite genus p, p greater than unity. They are to be without singularities. More specifically we will suppose that the points in the neighborhood of any point of the surface can be put into one to one continuous correspondence with the points in the neighborhood of some point in a plane, in such a manner, that for the neighborhoods considered, the cartesian coördinates, x, y, and z, of a point of the surface, be continuous functions of the cartesian coördinates, u, v, of the plane, provided with continuous partial derivatives up to the fourth order, while further

$$\left[\frac{D(xy)}{D(uv)}\right]^2+\left[\frac{D(xz)}{D(uv)}\right]^2+\left[\frac{D(yz)}{D(uv)}\right]^2 \neq 0.$$

The given surface will eventually be mapped upon the interior of a unit circle, in a manner that will be one to one and continuous as far as the neighborhoods of any two corresponding points are concerned, but by a correspondence that will be one to infinity as a whole. With that end in view there will now be defined a special group of linear transformations of a complex variable carrying the interior of the unit circle into itself. The method of definition will be similar to the method used by Poincaré in the article cited below.* From the "fundamental domain" of this group will be formed a canonical surface which can be mapped on the interior of the unit circle in the desired manner, and with the aid of which the given surface can also be so mapped.

3. **The polygon S_0.** Let there be given a unit circle. In the plane of this circle let there be drawn a second circle concentric with the unit circle and with a radius r greater than one. Let p be a positive integer greater than one. On the circle of radius r let there be placed $4p$ equidistant points. With these points as centers let there be drawn $4p$ circles orthogonal to the unit circle. Now let r and the $4p$ circles vary, the $4p$ circles still remaining orthogonal to the unit circle and still having their centers equidistant on the circle of radius r. As r becomes arbitrarily large the $4p$ circles will approach straight lines passing through the center of the unit circle, and such that each makes an angle of $\pi - 2\pi/4p$ with its successor or predecessor. For very large values of r there will thus be formed a curvilinear polygon, lying

* *Théorie des groupes fuchsiens*, Acta Mathematica, vol. 1 (1882), p. 1.

within the unit circle, containing the center of the unit circle, and absolutely symmetrical with respect to the center of the unit circle (cf. Fig. 1). If r now decrease from very large values to small values, for some value of r the $4p$ circles will become tangent to each other. That is, the interior angles of the curvilinear polygon will diminish from $\pi - 2\pi/4p$ to zero. But for $p > 1$

$$\pi - \frac{2\pi}{4p} > \frac{2\pi}{4p} = \frac{\pi}{2p}$$

so that for a properly chosen value of r the curvilinear polygon will have interior angles of the magnitude of $\pi/2p$. The sides of this polygon will all be segments of circles orthogonal to the unit circle and will be equal in length. The polygon will contain the center of the unit circle and be absolutely symmetrical with respect to that center. The sum of the interior angles of the polygon will be 2π. This polygon will be denoted by S_0 and the corresponding value of r by r_0. (Cf. Fig. 1, for the case $p = 2$.)

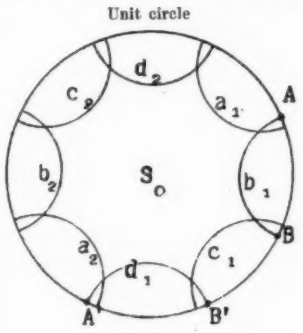

Fig. 1.

A property of S_0 to be used later is that *among the $4p$ circles bounding S_0, alternating circles do not meet.* Suppose they did meet. If r and the common radius of the $4p$ circles should then be increased, the $4p$ circles still remaining orthogonal to the unit circle, the alternating circles would still continue to meet. Now the interior angles of S_0 have a magnitude $\pi/2p$ which is less than $\pi/2$ for $p > 1$. On the other hand if r should be increased beyond all limit the interior angles of the polygon thereby formed, say S', would approach $\pi - 2\pi/4p$ which is greater than $\pi/2$ for $p > 1$. Thus for some value of $r > r_0$ the interior angles of S' would be right angles. Thus of any three successive circles bounding S', the first and last would be orthogonal to the second and at the same time would be orthogonal to the unit circle. This is impossible. For two circles orthogonal to two non-tangent circles which meet, cannot themselves meet.

4. **A group with principal circle.** The polygon S_0 will be used to define a "group with principal circle."* The $4p$ sides of S_0 will be taken in one of their two circular orders and labelled as follows:

$$a_1\, b_1\, c_1\, d_1\, a_2\, b_2\, c_2\, d_2, \cdots, a_p\, b_p\, c_p\, d_p.$$

* H. Poincaré, *Théorie des groupes fuchsiens*, loc. cit. Also L. R. Ford, *An introduction to the theory of automorphic functions*, Edinburgh Mathematical Tracts.

No confusion need arise if the circles bearing these sides be indicated by the same letters. We will term a_k conjugate to c_k, and b_k conjugate to d_k, where k ranges over the integers from 1 to p. Now a reflection of the plane in the radical axis of a_k and c_k will carry S_0 into itself and carry a_k into its conjugate side c_k. On the other hand a reflection of the plane in c_k will carry c_k into itself and S_0 into an adjacent polygon bordering S_0 along c_k. The product of these two reflections will be a transformation expressible as a linear transformation of a complex variable, and will carry the side a_k into the side c_k, and S_0 into a polygon bordering S_0 along c_k. Similarly there exist linear transformations of a complex variable which will carry any one of the sides of S_0 into its conjugate, and S_0 into a polygon bordering S_0 along the second of these conjugate sides. These transformations and all possible combinations of them will form a group G. All transformations of this group will carry the unit circle into itself. Points and regions which are the images of each other under this group will be termed *congruent*.

It is here necessary to make a convention, that just one side of each pair of conjugate sides of S_0 shall be considered as belonging to S_0, and just one of S_0's vertices. With this understood it may be stated that the set of all curvilinear polygons congruent to S_0 cover the interior of the unit circle once and only once. About each of the vertices of each of the polygons there will be grouped $4p$ polygons each with one corner at the given vertex. Each of these corner regions at such vertices will be congruent to just one corner in S_0, and conversely each corner of S_0 will be congruent to just one corner at each such vertex. For the proofs of the results of this paragraph the reader is referred to the literature on the subject.

Nature of the fundamental group

5. **Distances between congruent points.** The circles orthogonal to the unit circle may be looked upon as representing the straight lines of a hyperbolic non-euclidean (written N E) geometry in which the points of the hyperbolic plane consist of the points interior to the unit circle. If we suppose the unit circle has its center at the origin of the euclidean plane, the N E length of any rectifiable curve interior to the unit circle is taken as the value of the integral

$$\int \frac{2\,ds}{1-x^2-y^2},$$

taken along the given curve, and the N E distance between any two points within the unit circle as the N E length of the N E straight line joining the two given points. It is possible to prove directly that the N E length and

distance as defined by the above integral remain unchanged under any linear transformation of the complex variable that carries the interior of the unit circle into itself.

The NE *distance between any two congruent points has a lower limit not zero.*

It follows from the explicit nature (cf. § 4) of the transformations which carry S_0 into the polygons which are contiguous to S_0, that in a sufficiently small neighborhood of any point of S_0, or of its boundary, there are no pairs of mutually congruent points. Now if the lemma were false, it would be possible to pick out an enumerable infinity of pairs of congruent points whose NE distances would approach zero as the enumerating integer n became infinite. Each of these pairs could be replaced by a congruent pair of which one member at least might be taken so as to lie in S_0 or on its boundary. NE distances between the two points of any pair would be the same as before. These points in S_0 would have at least one limit point in S_0 or on its boundary. In every neighborhood of this limit point there could be found points which would be mutually congruent. This, as just stated, is impossible and the lemma is proved.

6. **Nature of the group.** *The group can contain no elliptic transformations.* An elliptic transformation carrying the unit circle into itself would have one fixed point within the unit circle and one without. Now there are in every neighborhood of a fixed point of an elliptic transformation points which are mutually congruent under the transformation. This is here impossible according to the result of the preceding section.

The group can contain no parabolic transformations.

To establish this fact it will be convenient to carry the interior of the unit circle into the upper half plane by means of a linear transformation of the complex variable. The integral

$$\int \frac{2\,ds}{1-x^2-y^2}$$

will go over into the integral

$$\int \frac{ds}{y}.$$

If now the points above the axis of reals be considered as the points of a hyperbolic NE plane, the circles orthogonal to the axis of reals as the straight lines of this geometry, and the value of the integral $\int (1/y)\,ds$ taken along any curve segment in the upper half plane as the NE length of that curve segment, then any curve segment within the unit circle may be considered as

unchanged in N E length by the transformation from the unit circle to the upper half plane.

Corresponding to the fundamental group carrying the unit circle into itself there will exist a transformed group carrying the upper half plane into itself. If the first group had a parabolic transformation the transformed group would have one also. Now any parabolic transformation, say T, which carried the axis of reals into itself would have its fixed point on the axis of reals. Suppose now the upper half plane be carried into itself and thereby the group be still further transformed in such a manner that the fixed point of the parabolic transformation T goes into the point at infinity. N E distances would thereby be unchanged but the original parabolic transformation would now be transformed into a translation parallel to the axis of reals through a euclidean distance, say d. A point, no matter how remote from the axis of reals, would still be carried by this parabolic transformation through this euclidean distance d. If sufficiently remote, however, from the axis of reals, the N E distance between two such congruent points is seen with the aid of the integral

$$\int \frac{ds}{y}$$

to be arbitrarily small. This is contrary to the result of the preceding section. Thus the group can have no parabolic transformations.

Now the only linear transformations of a complex variable which carry the interior of a circle into itself are either elliptic, parabolic, or hyperbolic. *Hence all the transformations of the group are of the hyperbolic type.*

7. A lemma on polygons near the unit circle. The N E distance between any two points of S_0 has an upper limit. The same upper limit necessarily holds for distances between any two points on any one of the congruent polygons. Accordingly a polygon which has any point sufficiently near the unit circle will be arbitrarily small in its euclidean dimensions, for otherwise its N E dimensions as measured by the N E integral

$$\int \frac{2\,ds}{1-x^2-y^2}$$

would become arbitrarily large. *The polygons accordingly cluster about each point of the unit circle.*

The following lemma will now be proved.

LEMMA. *If a polygon lie sufficiently near the unit circle, all of its bounding circles, with the possible exception of one of them, will be arbitrarily small.*

Denote by O the center of the unit circle. Let A and B be two points on the unit circle at the extremities of a diameter. Let d be a small positive constant. Consider all the radii through O, except those that make an angle with OA numerically less than d. Now let O move along AB towards B. At the same time let each of the radii not excepted be replaced by an arc of a circle orthogonal to the unit circle, and drawn from O to the unit circle so as to make the same angle with OA as the radius to be replaced. Denote this pencil of circles through O by C. If O move along AB to a point sufficiently near B, it is obvious that the circles of the pencil C just described will become arbitrarily small.

To prove the lemma consider the $8p$ radii which pass from the center O to the points on the unit circle where the circles bounding the polygon S_0 (cf. § 3) meet the unit circle. Denote this set of radii by R. Let d, the constant of the preceding paragraph, here be a positive quantity less than half the magnitude of the angle between any two successive radii of the set R. From the result of the preceding paragraph it follows that if O and S_0 be carried by a transformation of the group respectively into a point O' and a new polygon sufficiently near the unit circle, then the set of circular segments, say C', into which the radii of the set R are transformed will all be arbitrarily small *except* any of them which make an angle with $O'O$ of magnitude less than d. But from the choice of d there will be at most one circle of C' making an angle with $O'O$ of magnitude less than d. Now the circles bounding the transformed polygon and the circles of the set C' will intersect the unit circle in the same points, and accordingly, with the possible exception of one of them, each of the circles bounding the transformed polygon will meet the unit circle in two arbitrarily near points, and being orthogonal to the unit circle will consequently be arbitrarily small.

8. **Fixed points of the transformations of the group.** The distribution of the fixed points of the transformations of the group will now be considered.

(*a*). Let T be a transformation of the group. T is hyperbolic, and accordingly has two fixed points on the unit circle. Let A and B be two points on the unit circle not the fixed points of T, and let A' and B' be respectively the images of A and B under T. It follows from the nature of a hyperbolic transformation that in case the fixed points of T separate A and B, $ABA'B'$ will appear on the unit circle *in the circular order* $ABB'A'$, and that in the contrary case they will appear *in the order* $ABA'B'$ or else *in the order* $AA'BB'$.

(*b*). Consider any two conjugate circles of the circles bounding the polygon S_0. Suppose the first of these two circles meets the unit circle in the points A and B and the second of the two circles meets the unit circle in the points A' and B'

where A is congruent to A' and B to B' (cf. Fig. 1). According to the result of § 3 no two conjugate circles intersect. A reference to the explicit description of the transformations which carry any circle bounding S_0 into its conjugate (cf. § 4) will now show that $ABB'A'$ *appear on the unit circle in the order* $ABB'A'$. After subjection to any other transformation of the group the circular order of these four points will still be the same.

With the aid of the results of the preceding paragraphs (a) and (b) the following lemma can be proved. This lemma will be used to show that among the geodesics considered the types that correspond to closed geodesics are everywhere dense in a sense to be explained later.

LEMMA. *There exists a transformation of the group which has fixed points arbitrarily near the end points of any preassigned arc of the unit circle.*

Let M_1 and M_2 be the end points of a preassigned arc of the unit circle. Let e be an arbitrarily small positive constant less than half the distance between M_1 and M_2. According to the lemma of § 7, it is possible to choose a polygon S_1 so near M_1 that all of its bounding circles with the possible exception of one of them will lie within e of M_1. Let S_2 be a second polygon similarly chosen relative to M_2. Since $p > 1$ and the number of pairs of conjugate sides of a polygon is $2p$, S_1 and S_2 each have at least four pairs of conjugate sides, and on both S_1 and S_2 at least three of these pairs of conjugate sides lie entirely within e of M_1 and M_2 respectively. It is accordingly possible to choose a side of S_1 such that its intersections with the unit circle, say A_1 and B_1, and the intersections of its conjugate side with the unit circle, say A_1' and B_1', lie within e of M_1, while at the same time the four points $A_2 B_2 A_2' B_2'$ corresponding to $A_1 B_1 A_1' B_1'$ under the transformation of the group that carries S_1 into S_2 also lie within e of M_2.

According to the choice of e, the e neighborhood of M_1 is entirely distinct from the e neighborhood of M_2. Accordingly $A_1 B_1 A_2 B_2$ have either the circular order $A_1 B_1 B_2 A_2$ or else the circular order $A_1 B_1 A_2 B_2$. In the first case it follows from the result of paragraph (a) that the transformation of the group that carries $A_1 B_1$ into $A_2 B_2$ will have one of its fixed points on the arbitrarily small arc between A_1 and B_1 and the other fixed point on the small arc between A_2 and B_2, and the theorem is proved for that case.

Consider the second case. Here $A_1 B_1 A_2 B_2$ appear in the order $A_1 B_1 A_2 B_2$. According to the result of paragraph (b) $A_1 B_1 B_1' A_1'$ always appear in the circular order $A_1 B_1 B_1' A_1'$. It follows that in this second case $A_2 B_2 B_1' A_1'$ appear in the order $A_2 B_2 B_1' A_1'$. Using again the result of paragraph (a), the transformation of the group which carries $A_2 B_2$ into $A_1' B_1'$ is seen to be one which has one fixed point on the small arc between A_2 and B_2 and the other fixed point on the small arc between A_1' and B_1', and the lemma is proved for the second case as well.

THE GIVEN SURFACE MAPPED ON THE UNIT CIRCLE

9. **Definition of the correspondence.** Let a closed surface be formed from S_0 by continuously deforming S_0 so as to bring congruent points of conjugate sides into coincidence. S_0 has $2p$ pairs of conjugate sides. Hence the resulting closed surface, say T, will be of genus p. Each point P of T will arise from one point P_0 of S_0. The correspondence between T and S_0 can be extended over the interior of the unit circle by requiring that P shall not only correspond to P_0, but also to all points of the unit circle congruent to P_0. The correspondence thereby established between the interior of the unit circle and T is everywhere continuous and one to one as far as the neighborhoods of corresponding points are concerned, but one to infinity as a whole.

Now by a fundamental theorem of analysis situs any two closed two-sided surfaces of the same genus can be put into one to one continuous correspondence. The original surface, say Σ (cf. § 2), can then be put into one to one continuous correspondence with that one of the above surfaces T which has the same genus as does Σ. The surface Σ will thereby be mapped on the interior of the unit circle in a manner essentially the same as the manner in which T has just been mapped upon the interior of the unit circle.

10. **Surface distances and NE distances.** In contradistinction to NE lengths and distances, surface lengths and distances will now be defined.

DEFINITION. If a curve segment h within the unit circle corresponds on the original surface to a curve segment k that is rectifiable, the ordinary length of k will be termed the *surface length of* h. Before defining the surface distance between two points within the unit circle the following lemma is needed. *Of the curves within the unit circle joining any two non-coincident points, there is at least one which corresponds on the original surface to a geodesic segment whose surface length is at least as small as that of any other curve joining the two given points.* This lemma is essentially equivalent for the case of geodesics to Hilbert's *a priori* existence theorem of the calculus of variations.* A method which can be used without any essential change to prove this lemma is given by G. D. Birkhoff† for a similar problem in the article cited below.

DEFINITION. The lower limit of the surface lengths of curves joining two given points within the unit circle will be termed the *surface distance* between these two points. The surface distance between two given points will be denoted by D_g while the NE distance between the same two points will be denoted by D_n.

* Cf. Bolza, *Vorlesungen über Variationsrechnung*, 1909, p. 428.

† *Dynamical systems with two degrees of freedom*, these Transactions, vol. 18 (1917), § 9, p. 219.

LEMMA 1. *Corresponding to a positive constant a there exists a positive constant b such that any pair of points interior to the unit circle for which*

$$D_g \leqq a$$

will be such that

$$D_n < b.$$

Any pair of points whose surface distance is at most a can be replaced by a congruent pair of which one member lies on S_0 and the other on the closed region consisting of all points whose surface distance from points of S_0 and its boundary is at most a. These new pairs will moreover have the same surface distances as the original pairs. But the N E distances between any two points lying on a closed region are all less than a properly chosen constant which we may take for b. The lemma thus is proved.

LEMMA 2. *There exist constants A and m such that for any pair of points interior to the unit circle, if either*

$$D_n > A,$$

or

$$D_g > A,$$

then

$$D_n < m D_g.$$

Let there be given two points whose surface distance is D_g. Starting from one end of a geodesic of surface length D_g joining the two given points, let there be marked off segments of surface length equal to the constant a of Lemma 1. The line will thereby be divided into not more than

$$\frac{D_g}{a} + 1$$

parts. According to the preceding lemma there is between the end points of each of these parts an N E distance less than the constant b. The N E distance between the end points of the geodesic is accordingly less than

$$b\left[\frac{D_g}{a} + 1\right].$$

That is,

$$D_n < \frac{b}{a} \cdot D_g + b. \tag{1}$$

If then

$$D_g > b \tag{2}$$

(1) becomes

$$D_n < \frac{b}{a} \cdot D_g + D_g$$

or

$$D_n < \left(\frac{b}{a} + 1\right) D_g. \tag{3}$$

Setting

$$\frac{b}{a} + 1 = m$$

we have

$$D_n < m \cdot D_g \tag{4}$$

subject to the condition (2). Now (1) holds unconditionally, and shows that D_g becomes infinite with D_n, so that the condition (2) will be fulfilled if for a sufficiently large constant c

$$D_n > c. \tag{5}$$

As seen now (4) holds if either (2) or (5) holds. Taking A as the larger of the constants b and c, and m as the constant m of (4) the inequalities of the lemma accordingly follow.

It may be observed that there is a large degree of reciprocity between D_n and D_g. In fact the two preceding lemmas and their proofs will remain true if the terms N E distance and surface distance be interchanged together with D_n and D_g. Lemma 2 and the lemma obtained by interchanging D_n and D_g in Lemma 2 may be combined into the following fundamental lemma:

LEMMA 3. *If either the* N E *distance or the surface distance between two points exceeds a properly chosen constant, then the* N E *distance lies between two positive constant multiples of the surface distance, and the surface distance lies between two positive constant multiples of the* N E *distance, where the constants involved are fixed once and for all for the whole unit circle and surface.*

11. **Proof of Lemma 7.** The question arises, do there exist geodesic segments which recede arbitrarily far from the N E straight lines joining their end points and which still have a surface length as small as that of any other geodesics joining their end points? This question is answered in Lemma 8. This section is devoted to proving Lemma 7 which is the principal aid in proving Lemma 8.

A slight modification of the proof of Lemma 2 will serve to establish the following lemma. The constants b and m of the following lemma are the constants b and m used in Lemmas 1 and 2.

LEMMA 4. *There exist constants b and m such that, if a geodesic segment has a surface length g for which $g > b$, then the end points of this segment can be joined by a curve all of whose points lie within an* NE *distance b of the given geodesic segment, and whose* NE *length, v, is such that*

$$v < mg.$$

LEMMA 5. *Corresponding to a positive constant μ there exists a positive constant R so large that if all the points of a curve segment h are at an* NE *distance R from some* NE *straight line, L, and the end points of h project through* NE *perpendiculars into the end points of a segment of L, say k, of more than unit length, then h has an* NE *length exceeding μ times the* NE *distance between its end points.*

The locus of points at a constant NE distance from L is a circle meeting the unit circle in the same points as does L. The segment h lies on this circle.

To simplify the proof let the interior of the unit circle be carried into the upper half plane by a linear transformation which carries the end points of L respectively into the origin, O, and the point at infinity. This can be done, as in § 6, without altering any NE distances. L will go into an ordinary straight line, L', perpendicular to the axis of reals at the origin (see Fig. 2). The transform of k, say k', will lie on this line. The circle on which h lies will go into an ordinary straight line through the origin making an angle with the axis of reals which we denote by α, measuring the angle from the axis of reals in the counterclockwise direction. We may suppose further that $\alpha < \pi/2$, since that could be brought about without altering any NE distances by a reflection of the upper half plane in L'. The transform of h, say h', will lie on this line, making an angle α with the axis of reals. The NE perpendiculars from the end points of h to those of k will go respectively into two arcs, say a' and b', of two circles both with center at the origin and leading from the end points of h' to those of k'. To compute the NE length of a', b', h', and k', suppose the

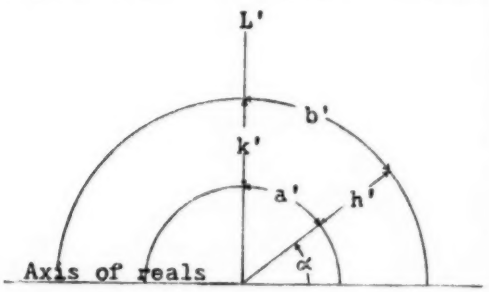

Fig. 2.

plane referred to polar coördinates ϱ and φ with pole at the origin. Denote by ϱ_1 the value of ϱ at the lower end point of k', and by ϱ_2 its value at the upper end point. The polar coördinates of the end points of h' will be $(\varrho_1\ \alpha)$ and $(\varrho_2\ \alpha)$. The N E length of k' as evaluated through the integral $\int (ds/y)$ is

$$\int_{\rho_1}^{\rho_2} \frac{d\varrho}{\varrho} = \log \frac{\varrho_2}{\varrho_1}. \tag{1}$$

The N E length of h', along which φ equals the constant α, is

$$\int_{\rho_1}^{\rho_2} \frac{d\varrho}{\varrho \sin \alpha} = \log \frac{\varrho_2}{\varrho_1} \cdot \frac{1}{\sin \alpha}. \tag{2}$$

The N E lengths of a' and b' are the same, namely

$$\int_{\alpha}^{\frac{\pi}{2}} \frac{\varrho_1\, d\varphi}{\varrho_1 \sin \varphi} = \log \cot \frac{\alpha}{2}, \qquad 0 < \alpha < \frac{\pi}{2}. \tag{3}$$

It is desired to show that if the end points of h' be joined by the broken line $a'\ k'\ b'$, there results a curve whose N E length has a ratio to the N E length of h' that approaches zero as R becomes infinite. The ratio of the N E length of the broken line $a'\ k'\ b'$ to that of h' is

$$\frac{2 \log \cot \frac{\alpha}{2} + \log \frac{\varrho_2}{\varrho_1}}{\log \frac{\varrho_2}{\varrho_1}} \sin \alpha = \left[\frac{2 \log \cot \frac{\alpha}{2}}{\log \frac{\varrho_2}{\varrho_1}} + 1 \right] \sin \alpha.$$

Now by hypothesis the N E length of k', namely $\log (\varrho_2/\varrho_1)$, is greater than one, so that the above ratio is less than

$$\left(2 \log \cot \frac{\alpha}{2} + 1\right) \cdot \sin \alpha. \tag{4}$$

R, the length of a' or b', is given by (3) and becomes infinite only if α approaches zero. But as α approaches zero the expression (4) approaches

zero, so that for R sufficiently large the N E length of h' will certainly exceed μ times the N E length of the broken line $a'k'b'$ joining the end points of h', and hence the N E length of h' will exceed μ times the shortest N E distance between the end points of h', as was to be proved.

DEFINITION. Let k be any curve segment within the unit circle. *The* N E *distance* between the end points of k will be denoted by $D_n(k)$, and the *surface distance* between the end points of k will be denoted by $D_g(k)$.

LEMMA 6. *Corresponding to positive constants b and m there exists a positive constant r so large that if a curve segment, v, has its end points at an* N E *distance r from some* N E *straight line L, if further all of v's points are at an* N E *distance from L exceeding $r-b$, and if v projects into a segment of L whose* N E *length exceeds unity, then the* N E *length of v exceeds m times the* N E *distance between v's end points.* (See Fig. 3.)

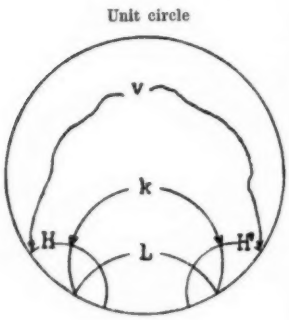

Fig. 3.

Denote by k a curve segment projecting through N E perpendiculars into the same segment on L as does v, and consisting further of points all at an N E distance $r-b$ from L. Let the N E length of v and k be denoted also by v and k respectively. We seek to prove that, for r sufficiently large,

$$v > m \cdot D_n(v). \tag{1}$$

To that end we shall obtain inequalities connecting v with k, k with $D_n(k)$, $D_n(k)$ with $D_n(v)$, and by combining these inequalities obtain the inequality (1).

By the aid of the construction of the preceding proof it is seen that

$$v > k.$$

Taking the constant μ of the preceding lemma as here equal to $m+1$ it follows from that lemma that if $r-b$, and hence r, be sufficiently large,

$$k > D_n(k) \cdot (m+1).$$

The two ends of v project into the two ends of k through N E perpendiculars to L, say H and H' respectively. Now the end points of v can be joined through the mediation of k by joining the first end point of v to the first end point of k, tracing out that segment of H that lies between these two

points, following this segment by tracing out the N E line joining the end points of k and proceeding finally to the second end point of v by tracing out that portion of H' which lies between the second end points of v and k. In accordance with one of the hypotheses of this lemma the portions of H and H' thereby traced out are each of N E length b, so that the whole broken curve thus traced out will have an N E length $2b + D_n(k)$. This length will be at the least $D_n(v)$, since $D_n(v)$ stands for the shortest N E distance between the end points of v. Thus

$$2b + D_n(k) \geqq D_n(v) \text{ or } D_n(k) \geqq D_n(v) - 2b.$$

Combining the preceding inequalities we have

$$(2) \qquad v > (m+1)[D_n(v) - 2b] = (m+1) D_n(v) - (m+1) 2b.$$

Now as r becomes infinite $D_n(v)$ will do likewise, and, for all cases in which the segment v projects on L into a segment of at least unit N E length, $D_n(v)$ will become infinite uniformly with r. Suppose r so large therefore that

$$(3) \qquad D_n(v) > (m+1) 2b.$$

Adding (2) and (3) we have

$$v > m \cdot D_n(v)$$

with the condition simply that r be sufficiently large. The proof is accordingly complete.

LEMMA 7. *There exists a constant r such that if all the points of a segment of a geodesic g are at an* N E *distance at least r from some* N E *straight line L, if further the end points of the geodesic segment g are exactly at an* N E *distance r from L, and if g projects in the* N E *sense into a segment of L of more than unit* N E *length, then the surface length g of the given geodesic exceeds the surface distance between its end points.*

It is required to prove that for r sufficiently large

$$(1) \qquad g > D_g(g).$$

Now as r becomes infinite $D_n(g)$ will also become infinite. It follows from Lemma 3 that $D_g(g)$ will do likewise, so that according to Lemma 4 there exist positive constants b and $\bar{m}$ such that if r exceed a properly chosen

positive constant the end points of g can be joined by a curve v whose points are each within an N E distance b of g and whose N E length v is such that

$$\overline{m}\, g > v.$$

We now have a relation between g and v. We shall obtain a relation between v and $D_n(v)$, and a relation between $D_n(v)$ and $D_g(v)$. Combining these relations (1) will be obtained.

If r exceed a second constant so large that $D_n(v)$ exceeds the constant required for the application of Lemma 3, we will have

$$D_n(v) > A\, D_g(v),$$

where A is a positive constant. If further r exceed the positive constant required for the application of Lemma 6 for the case where the m of Lemma 6 equals $\overline{m}/A$, then in addition to the preceding inequalities there will also result the inequality

$$v > \frac{\overline{m}}{A}\, D_n(v).$$

If the above inequalities be combined there results the inequality

$$g > D_g(v),$$

subject simply to the condition that r exceed a properly chosen constant.

A FUNDAMENTAL CLASS OF GEODESICS

12. Geodesics of Class A. DEFINITION. A geodesic segment h will be said to be of *Class* A, if, on the original surface, h is at least as short as any other rectifiable curve joining h's end points and capable of being continously deformed on the surface into h without moving its end points. An image of a geodesic segment of Class A on the hyperbolic plane for simplicity also will be termed a geodesic segment of Class A. The definition of geodesic segments of Class A reduces in the hyperbolic plane to the following. An image of a geodesic segment within the unit circle is of Class A if its surface length is at least as small as that of any curve joining its end points and possessing a surface length. *An unending geodesic* will be said to be of Class A if each of its finite segments is of Class A.

LEMMA 8. *There exists a positive constant R, so large, that within the unit circle, no segment of a geodesic of Class* A *can recede an* NE *distance greater than R from the* NE *straight line joining its end points.*

Denote the given geodesic segment by g, and the NE straight line joining its end points by L. Consider the constant r whose existence is affirmed in Lemma 7. If g has points at more than r NE units from L, let P' be such a point. Let P and Q be the first points following and preceding P', respectively, that are at an NE distance r from L. If the segment PQ projects through NE perpendiculars into a segment on L of more than unit NE length, it follows from Lemma 7 that the surface length of the segment PQ exceeds the surface distance between P and Q contrary to the hypothesis that the given geodesic segment is of Class A.

If on the other hand the segment PQ projects into a segment on L of unit NE length or less than unit NE length, while P and Q are at an NE distance r from L, the NE distance between P and Q will be at most $2r+1$, and hence the surface distance between P and Q will in all cases be inferior to some positive constant, say d. Since g is of Class A the surface distance between P' and P is at most that between P and Q, or at most d. It follows from Lemma 3 that the NE distance of P' from P is always less than some constant, say q. Hence P' is in all cases at most at NE distance $r+q$ from L, and the theorem is proved.

13. **Unending curves of the same type.** A definition will presently be given of what is meant by two unending curves being of the same type. This definition will be an extension of the definition of what is meant by two closed curves being mutually deformable. We begin accordingly by considering closed curves. A curve continuously deformable into a point either on the given surface or on the unit circle, will have an image deformable into a point on the other. A closed curve, say h, not deformable into a point on the given surface, if traced out just once will have as images on the unit circle an infinite number of curve segments whose end points are congruent, non-coincident points. Let A and B be respectively the end points of one of these images.

A is congruent to B under a certain transformation T. The set of all points within the unit circle congruent to A under T and its inverse all lie on a circle through the fixed points of T and cluster in the neighborhood of these fixed points. The set of all segments congruent to AB under T and its inverse join up these points congruent to A so as to form a single curve passing from one fixed point of T to the other. An unending curve such as this, consisting of an unending succession of segments all congruent under the same transformation of the group, will be called a *periodic curve*. Traced out in either sense such a curve corresponds on the original surface to a closed curve traced out an infinite number of times.

It follows from the fact that the correspondence between the surface and unit circle is one to one and continuous as far as the neighborhood of corresponding points is concerned, that if the given closed curve, h, be continuously deformed on the surface, then the image AB will vary continuously through corresponding points, A and B always remaining congruent under the same transformation. At the same time the variation of the complete set of points congruent to A under T and its inverse is one in which the circle on which these points lie turns about the fixed points of T while the points themselves vary continuously on the circle in such a manner as to remain always congruent under T. Thus two closed curves which on the original surface are not deformable into points, but which are mutually deformable into each other, will correspond within the unit circle, among other images, to pairs of periodic curves with end points at the same points on the unit circle. For the sake of completeness the following is added. Two curves which correspond to two periodic curves within the unit circle with the same end points on the unit circle and with successive portions congruent under the *same* transformation of the fundamental group, are mutually deformable on the original surface.

DEFINITION. Two unending curves within the unit circle will be said to be of the *same type* if there exists a positive constant C such that every point of either curve has an N E distance from some point of the other less than C. Two unending curves *on the original surface* will be said to be of the *same type* if among their respective images on the unit circle there are at least two curves of the same type.

Two circular arcs within the unit circle are of the same type if their end points on the unit circle are the same, and of different types if either of their end points are different. Every circular arc with two end points on the unit circle is thus of the type of the N E straight line with the same two end points on the unit circle. *Only those curves will be considered which are of the types of* N E *straight lines.* There is thereby included a type for every periodic curve, that is, curves which correspond on the original surface to closed curves. Future developments will still further show the inclusiveness of types of curves as restricted to N E straight lines. From Lemma 3, it follows that N E distance and surface distances become infinite together uniformly. Accordingly if two unending curves are of the same type there exists a constant D such that every point of either curve has a surface distance less than D from some point on the other curve. It is not difficult to show that if on the original surface two curves are closed and mutually deformable they have for their images on the unit circle an infinite number of pairs of periodic curves of the same type. It is in this sense that *the property of two curves being of the same type is an extension of the property of two curves being*

mutually deformable. The definition of what is meant by two curves being of the same type will be given in the following paragraph more in the spirit of analysis situs. The statements of the following paragraph will not be proved as they will not be used later.

Let the given surface, $p > 1$, be cut so as to form a single simply-connected piece S. Let there be provided an enumerable infinity of copies of S each spread over the original surface in exactly the position in which they were before the surface was cut. Without altering the position of any of these copies of S, these copies of S can be successively joined in a manner analogous to the way in which the congruent polygons of the unit circle were joined, so as to "heal" up all of the cuts and to form an infinitely-sheeted, simply-connected, unbounded surface, say M, the neighborhood of each point of which will consist on M of an exact copy of the neighborhood of the point which is overhung on the original uncut surface. Two curves of M could now be defined as being of the same type if there exists a positive integer n, so large, that any point of either curve could be joined to some point of the other curve by a curve which has points in at most n of the copies of S making up M. From this definition it could be shown that *two curves of the same type on the given surface will correspond, on any surface which corresponds to the given surface in a one to one continuous manner, to two curves again of the same type.*

14. **Existence of geodesics of Class A.** An image on the unit circle of a geodesic on the original surface will be spoken of as a geodesic on the unit circle. By an *element* is meant a point and a direction through the point. By an element on a geodesic on the surface is meant a point on the geodesic together with a direction tangent to the given geodesic at the given point. By an *element on a geodesic on the unit circle* is meant a point P within the unit circle on such a geodesic, together with the direction of a tangent to the original geodesic at the point on the original surface corresponding to P. In either case an element may be represented by the coördinates x, y of its initial point in the unit circle, together with the direction cosines a, b, c of the direction tangent to the original geodesic on the surface. The element may be considered as represented by a point $(xyabc)$ in five dimensions. A set of elements will be said to have a *limit element* $(x_0\, y_0\, a_0\, b_0\, c_0)$ if the corresponding points in five dimensions have a limit point $(x_0\, y_0\, a_0\, b_0\, c_0)$. With these conventions the following definition is given.

DEFINITION. A geodesic G will be said to be a *limit geodesic* of a set of geodesics M not containing G, if every element on G is a limit element of elements on geodesics of the set M.

An element will be said to be on an NE straight line in the unit circle if its initial point P is on the NE straight line, and if its direction is the direction of a tangent to the NE straight line at P.

LEMMA 9. *If a geodesic is a limit geodesic of a set of geodesics of Class* A, *it is itself of Class* A.

Let H be an image of the limit geodesic within the unit circle. If H were not of Class A, it follows from the definition of geodesics of Class A, that H has some segment, h, whose end points can be joined by a curve k whose surface length k is less than the surface length h of the segment h. Set

$$k = h - e.$$

On the other hand, if the given geodesic is a limit geodesic of geodesics of Class A, it follows, from the property of continuous variation of a finite geodesic segment with its initial point and direction, that there exists a geodesic segment m, of Class A, with end points within $e/4$ of those of h, and with a surface length m, such that

$$m > h - \frac{e}{2}.$$

Now the end points of m can be joined through the mediation of k by first passing from one end point of m to the nearer end point of k, passing thereby along a curve which can be taken so as to have a surface length less than $e/4$; thence passing along k to k's other end point, thereby passing over an additional surface length of $h - e$; finally passing from the second end point of k to the second end point of m along a curve taken so as to have a surface length less than $e/4$; in all passing over a curve of surface length less than $h - (e/2)$, that is, over a length less then that of m. This is impossible if m be of Class A. Thus the lemma is proved.

THEOREM 1. *Corresponding to any* NE *straight line, there exists at least one geodesic of Class* A *of the same type. Conversely, every geodesic of Class* A *is of the type of some one* NE *straight line.*

Let there be given an NE straight line L. Let

$$\cdots P_{-2}\ P_{-1}\ P_0\ P_1\ P_2 \cdots$$

be an unending succession of points on L which follow each other on L in the order of their subscripts, and which with increasing subscripts approach one end point of L on the unit circle, and with decreasing subscripts approach the other end point of L. Within the unit circle let P_{-n} be joined to P_n by a geodesic of Class A, say g_n. According to Lemma 8, g_n will have at least one point, say Q_n, within an NE distance R of P_0. Let E_n be the geodesic

element on g_n at Q_n. The elements E_n will have at least one limit element, E. Let G be the geodesic within the unit circle bearing the element E. G will be shown to be of Class A, and to be a curve of the type of the given N E straight line.

If any finite segment of G be given, it follows from the definition of G, that there can be found among the segments g_n, one which possesses a subsegment lying arbitrarily near the given segment of G, not only in position but also in direction and length. A repetition of the proof of the preceding lemma will suffice to show that the given segment of G is of Class A, as was to be proved.

Every point of the segments g_n lies within an N E distance R of some point of L (cf. Lemma 8). All the limit points of the points of the segments g_n, including the points of G, accordingly lie at most at an N E distance R from points of L. G thus lies in a region bounded by the two circles at an N E distance of R from L. Any two such circles and L meet the unit circle in the same two points. G passes from one of these points to the other. From this it is obvious that each point of L, also, lies at most at a distance R from some point of G, and G is thereby proved to be of the type of L.

To prove the converse proposition let G be a geodesic of Class A, within the unit circle, and let

$$\cdots P_{-2}\, P_{-1}\, P_0\, P_1\, P_2 \cdots$$

be a set of points appearing on G in the order of their subscripts, and so chosen that their surface distances from P_0 increase beyond all limit as n becomes infinite in absolute value. Denote by h_n an N E straight line joining P_n to P_{-n}. At least one point of h_n, say Q_n, will be at an N E distance from P_0 not greater than the constant R of Lemma 8. Denote by E_n, the element on h_n at the point Q_n. The elements E_n will have at least one limit element, say E. The N E straight line, L, through E, will be shown to be of the type of G.

Let P be any point on G. Let m be the smallest value of n such that all segments $P_{-n}\, P_n$ for which $n > m$ include the point P. For $n > m$ there is a point A_n on each h_n which is within an N E distance R (cf. Lemma 8) of the point P. But as n becomes infinite the points A_n will have at least one cluster point which, according to the definition of L, will lie on L. This cluster point will be at most at an N E distance R from the given point P of G. G thus lies within the band consisting of all points within an N E distance R of L. G and L will accordingly be of the same type, and the theorem is completely proved.

As just stated, a geodesic of Class A lies within a band consisting of all points within an N E distance R of the N E straight line of the same type. This constant R is the constant of Lemma 8, and is entirely independent of

the particular geodesic of Class A that is considered. Combining this fact with the fact given in Lemma 3, that N E distances and surface distances become infinite together uniformly for the whole unit circle, we have the following theorem:

THEOREM 2. *There exists a positive constant D, fixed for all types of geodesics of Class* A, *such that no geodesic of Class* A, *and* N E *straight line of the same type, can recede either a surface distance or an* N E *distance exceeding D from each other.*

15. Intersecting and asymptotic geodesics of Class A.

THEOREM 3. *Two geodesics of Class* A *can intersect at most once within the unit circle.*

If two such curves, G and H, intersected twice in two points, A and B, the portions, g and h, respectively, of G and H, between the points A and B, would have the same surface length. For if, for example, h were shorter than g, G could not be of Class A, as follows directly from the definition of curves of Class A. But g and h cannot have the same surface length. For consider a segment k of G large enough to contain the segment g as an interior segment. If the surface length of g equaled that of h, k would be unaltered in surface length if its subsegment g be replaced by the segment h. But the curve so obtained would have corners at A and B, and so could be shortened, contrary to the fact that k is of Class A, and hence could not be shortened. Thus the theorem is proved.

A geodesic of Class A within the unit circle passes from one end point of the N E straight line of the same type, to the other such end. Hence the theorem:

THEOREM 4. *If two* N E *straight lines have both end points distinct, the corresponding geodesics of Class* A *of the same type cross or do not cross each other within the unit circle, according to whether or not the given* N E *lines cross or do not cross each other.*

If two N E straight lines have one end point on the unit circle in common, the minimum N E distance from a point P on either to a point on the other approaches zero as the point P approaches the common end point. The N E straight lines are termed *asymptotic*. A direct result of Theorem 2 is accordingly

THEOREM 5. *If two geodesics of Class* A *are respectively of the types of two asymptotic* N E *straight lines, as they approach the common end point on the unit circle of the given* N E *straight lines, they come and remain within a finite surface distance of one another.*

THEOREM 6. *No two geodesics of Class* A *which are asymptotic can intersect within the unit circle.*

Let a and b be two geodesics of Class A which are asymptotic.

Choose as the positive senses of a and b the senses in which a and b are asymptotic. Suppose the theorem were false and that a and b did intersect in a point O (cf. Fig. 4). Now any broken geodesic which contained any one of the corners formed by a and b together with a segment of a or b on either side of O of at least unit surface length could be replaced by a curve joining its end points and with a surface length smaller by at least a positive constant e.

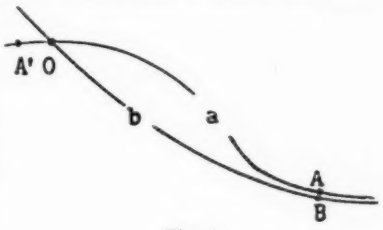

Fig. 4.

By hypothesis a and b are asymptotic. Accordingly it will be possible to find two points A and B, on a and b, so remote in the positive senses of a and b that they can be joined by a curve whose surface length AB is less than $e/4$:

$$AB < \frac{e}{4}.$$

Observe that the surface length of the segment AO of a cannot differ from that of BO on b by $e/4$ or more, for otherwise the larger of the two segments AO and BO, say AO, could be replaced by a broken curve of shorter surface length, namely ABO. Hence, in particular,

$$BO < AO + \frac{e}{4}.$$

Now let A' be any point on a, at least a unit surface distance beyond O in the negative sense. The segment AOA' of a can be replaced by a segment of shorter surface length. For adding to the two preceding inequalities the equality

$$OA' = OA',$$

we have

$$ABOA' < AOA' + \frac{e}{2}.$$

But the broken curve, $ABOA'$, contains a corner formed by a and b at O, so that it can be shortened by at least the constant e without altering its end points. The new curve will still join the end points of AOA', and will now in addition have a surface length less than that of AOA'. This is impossible if a is of Class A. The theorem thus is proved.

We have already defined what is meant by a limit geodesic of a class of geodesics (cf. § 14).

DEFINITION. We here define an NE straight line L to be a *limit line* of a set of NE straight lines M not including L, if every element of L is a limit element of elements on lines of the set M. A necessary and sufficient condition that L be a limit line of the set M is that there be in the set M lines whose end points on the unit circle lie, in the euclidean sense, arbitrarily near those of L.

THEOREM 7. *If a set M of* NE *straight lines have an* NE *straight line L for a limit line, the geodesics of Class* A *of the type of the lines of M will have at least one geodesic of the type of L as a limit geodesic.*

Let A be any element on L. It follows from the hypothesis of the theorem hat there exists a set of elements,

$$A_1 A_2 A_3 \cdots,$$

all on lines of M, and such that

$$\lim_{n \to \infty} A_n = A.$$

Let L_n be the NE straight line of M bearing A_n. Let C_n be a geodesic within the unit circle of Class A and of the type of L_n. According to Lemma 8, there exists on C_n at least one point, say P_n, within an NE distance R of the initial point of the element A_n. Let E_n be the geodesic element on C_n at P_n. The elements E_n will have at least one limit element, E. It follows from the result of Lemma 9, § 14, that the geodesic G, defined by E, is of Class A. It will now be proved that G is of the type of the NE straight line L.

It follows, from the property of continuous variation of a finite segment of a geodesic with its initial element, that every point P of G is a limit point, as n becomes infinite, of a properly chosen set of points on the geodesics C_n. But the latter points are each at an NE distance less than the constant R of Lemma 8 from some point of the corresponding NE straight line L_n. These points on the lines L_n will, however, as n becomes infinite, approach some point of L, say H, and this point H will be at most at an NE distance R from the given point P on G. Thus G is of the type of L as was to be proved.

THEOREM 8. *If there is more than one geodesic of Class* A *of a given type within the unit circle, there always exist two particular non-intersecting geodesics of Class* A *of the given type bounding a region outside of which no geodesics of Class* A *of the given type can pass.*

Let L be an NE straight line of the given type. Let M be a set of NE straight lines which have L as a limit line, which have no points in common

with L either within or on the unit circle, and which all lie on one side of L. Let N be a similar set of NE straight lines all on the other side of L. Let M' and N' be two sets of geodesics consisting of one geodesic of Class A of the type of each NE straight line of the sets M and N respectively. According to Theorem 7 the geodesics of the sets M' and N' each have at least one limit geodesic, say M_0 and N_0, respectively, that is of Class A, and of the type of L. The NE straight lines of the sets M and N do not cross L or have any end points in common with L. It follows from Theorem 4 that all geodesics of Class A of the type of L will lie in the region between any geodesic of the set M', and any second geodesic of the set N'. It follows further from the property of continuous variation of any finite segment of a geodesic with its initial element that the geodesics M_0 and N_0, by virtue of the fact that they are respectively limit geodesics of the sets M' and N', will, in case they are non-coincident, be non-intersecting, and have between them all geodesics, other than M_0 and N_0, of Class A and of the type of L.

Closed geodesics

16. **The variation of the distance of an arbitrary geodesic V of Class A from a given type of periodic curve h.** Let there be given an NE straight line, h, made up of successive segments all congruent under a transformation, T, of the fundamental group. We have previously termed h a periodic curve. Let k be a second NE straight line crossing h. Let K be the geodesic of Class A of the type of k. The transformation T carries h into itself. Accordingly the successive images of k,

$$\cdots k_{-2}\, k_{-1}\, k_0\, k_1\, k_2 \cdots,$$

under T and its inverse, will consist of a set of non-intersecting circles, all meeting h and shrinking down about the two ends of h, the fixed points of T. The successive images of K under T and its inverse, say

$$\cdots K_{-2}\, K_{-1}\, K_0\, K_1\, K_2 \cdots,$$

will accordingly not meet each other (cf. Theorem 4), and will, like the k_n, shrink down about the two fixed points of T. Now let there be given any NE straight line, v, and let V be a geodesic of Class A of the type of v. If v crosses any one of the k_n, V will cross the geodesic K_n with the same subscript.

Three cases are distinguished here. The end points of v on the unit circle may be distinct from those of the given periodic curve h. In this case v

crosses at most a finite number of the k_n. Or v may be asymptotic to h, in which case v will cross all of the k_n with subscripts exceeding a certain integer, or else less than a certain integer. Finally v may coincide with h, in which case v will cross all of the k_n. A study will now be made of the positions of the points of intersection of V with the different K_n.

Let A_0 be any point on K_0. Let A_n on K_n be the image of A_0 under T, or its inverse. N E distances will be measured along each K_n from the corresponding A_n. The sense that corresponds to positive distances on each of the K_n will be chosen so as to be unchanged through application of T. Now if V intersects K_n, denote this point of intersection by P_n and let s_n be the surface distance of P_n from A_n, measured along K_n. A proof will be given that the s_n behave in one of the following ways: (1) *remain constant,* (2) *increase with their subscripts,* (3) *decrease with their subscripts,* (4) *increase to a maximum and thereafter decrease, or* (5) *decrease to a minimum and thereafter increase. In the latter two cases the maximum or minimum may be taken on by a single s_n or by two successive s_n.* The proof will be given under the heads A, B, C, D, E.

A. *If three successive s_n are equal all of the s_n are equal.* An application of the transformation T to V will carry V into a new geodesic of Class A, say V'. If in particular $s_1 = s_2 = s_3$, P_1 and P_2 on V would go into P_2 and P_3 on V' so that V and V' would intersect in P_2 and P_3. According to Theorem 3, V must then be identical with V'. Thus each P_n will be carried by T into P_{n+1}, so that s_n equals s_{n+1}.

B. *If just two successive s_n are equal, the next preceding and the next following s_n, if they exist, are either both greater or both less than the equal s_n.*

Of the two regions into which V divides the unit circle, that region which a point enters upon being projected from V along any of the curves K_n in the positive sense of that particular K_n will be termed the region above V, and the other region the region below V. As before let V' be the image of V under the transformation T. Suppose now that $s_2 = s_3$ and that $s_1 > s_2$. To prove $s_3 < s_4$: V' will meet K_2 above V since $s_1 > s_2$. V' will meet K_3 in the same point as does V since $s_2 = s_3$, and at this point will pass from above to below V. Thereafter V' will remain below V, since it cannot cross V a second time. In particular V' will meet K_4 below V, so that $s_3 < s_4$ as was to be proved. Similarly it could be shown that if $s_2 = s_3$ and $s_1 < s_2$, then $s_3 > s_4$.

The preceding proof includes a proof of the following.

C. *If the s_n are equal for $n = m$ and $n = m-1$, V and V' meet on K_m.*

D. *If the s_n have a relative maximum or minimum for $n = m$, then between K_m and K_{m+1}, V and V' cross each other.* More explicitly suppose that

s_2 is greater than both s_1 and s_3. Then V' meets K_2 below V since $s_1 < s_2$, and meets K_3 above V since $s_2 > s_3$. Thus V' must cross V between K_2 and K_3. The case of a minimum among the s_n is similarly treated.

E. *The s_n cannot have more than one extremum of any sort.* For if the s_n had more than one extremum it follows from C and D that V and V' would have more than one point of intersection. V would then have to be identical with V'. All the s_n would be equal in that case.

A proof can now be given that the s_n behave only in the ways enumerated. Any sequence of numbers such as the s_n are either (1) all equal, or (2) increase or decrease monotonically with n, or (3) have at least one extremum. The first case is admittedly possible. The second case can be achieved, according to A and B, only when no two of the s_n are equal, and then is admittedly possible.

The third case, according to E, can occur only when there is a single extremum among the s_n. In this case the extremum will be taken on by not more than two successive s_n, as follows from A. Further, no two of the remaining s_n are equal, for if just two of them were equal, it would follow from B that they would constitute another extremum among the s_n, contrary to E; and if three or more of the remaining s_n were equal all of the s_n would be equal and we would have the first case.

Thus all three cases reduce to the cases originally stated to be possible.

17. **Periodic geodesics of Class A.** DEFINITION. Of the transformations of the fundamental group which have the same fixed points there are two which will be called the corresponding *primitive transformations.* They are inverses of each other. Each is such that every other transformation of the group with the same fixed points is either its multiple or the multiple of its inverse. Such transformation exists. For every transformation of the group which has the two given points for fixed points carries the NE straight line through these points into itself. With the aid of NE distances measured along this line the existence of the primitive transformations can be proved in exactly the same way as the existence of a primitive period is proved in the theory of simply periodic functions.

THEOREM 9. *A geodesic of Class A which is carried into itself by a transformation T' is also carried into itself by the corresponding primitive transformations.*

Denote the given geodesic by V. We shall make use of the notation of the preceding section, identifying the V and T of this section with the V and T of that section. Now according to the result of the preceding paragraph, there exists a positive integer m, such that

$$T' = T^m$$

where T is one of the two primitive transformations. Accordingly, upon applying T^m times to V, V will be carried into itself. Hence among the s_n used in the preceding section those whose subscripts differ by m must be equal. But if all of the s_n were not then equal, the s_n would have an infinite number of maxima and minima, which has been shown to be impossible. The s_n are then all equal. The result of applying T to V must carry V into itself. For the geodesic into which V is carried by T would otherwise intersect V in each of the points at which the s_n were equal; according to Theorem 3, this is impossible. Thus T carries V into itself, as was to be proved.

THEOREM 10. *A geodesic, V, of Class* A, *which if traced out in one of its senses, say the positive sense, comes and remains at an* NE *distance less than an assignable positive constant from a periodic* NE *straight line L, is either a periodic geodesic of the type of L, or else is asymptotic in its positive sense to a geodesic that is itself periodic, and is of the type of L, and of Class* A.

L is made up of successive portions congruent under a primitive transformation, say T. We will again make use of the notation of the preceding section identifying the V, T and L of the preceding section with the V, T and L of this theorem. We suppose that the subscripts n of the s_n of the preceding section are here so chosen that they increase as V is traced out in its positive sense. If the s_n are all equal, V must be periodic. The case where the s_n are not all equal will be considered now.

According to the result of the preceding section, if the s_n are not all equal, then for values of n exceeding a properly chosen integer, say m, the s_n must either all increase with n, or else all decrease with n. It follows from the hypothesis that V traced out in its positive sense comes and remains within an assignable positive NE distance of L, that the s_n are bounded, for $n > m$. Hence as n becomes infinite the quantities s_n approach a limit, say b.

Let k be any positive integer greater than m. The transformation which carries the geodesics K_k of the preceding section into K_0, carries L into itself, and V into a curve of Class A, say V_k, for which the corresponding s_n will be obtained from the s_n for V by advancing the subscripts by k. Thus the point of intersection of V_k with K_0, say Q_k, will be at an NE distance s_{0+k} from the point A_0 of K_0, from which NE distances along K_0 are being measured. As k becomes infinite the points Q_k will approach a point on K_0, say Q, whose surface distance from A_0 will equal $\lim_{k\to\infty} s_k = b$. Let E_k be the element on V_k, at the point Q_k. The elements E_k will have one or more limit elements. Let E be one such limit element. The initial point of E will be Q. The geodesic, say G, defined by E will be of Class A.

The point of intersection of G with any particular K_n, say K_3, for example, will be the limit of the points of intersection of the geodesics V_k with that

particular geodesic K_s, as k becomes infinite; so that G will intersect K_s at a surface distance from A_s on K_s equal to $\lim_{k\to\infty} s_{s+k} = b$. Similarly G will meet all other K_n at a surface distance from the corresponding A_n equal to b. It follows that the transformation T will carry G into itself, so that G is a periodic geodesic.

It remains to be proved that V is asymptotic to G. Now E was defined as a limit element of the elements E_k. But the elements E_k can have no other limit element than E. For if the elements had another limit element E', then E', like E, would serve to define a second periodic geodesic, say G', passing through the point Q. G' and G would intersect not only in Q but also in all the points congruent to Q under the transformation T. This is impossible for two geodesics of Class A. Hence $G' = G$, and $E' = E$, and the set E_k has no other limit element than E. This may be written

$$\lim_{k\to\infty} E_k = E. \tag{1}$$

Now E_k is the initial element of the portion of V_k, say h_k, between K_0 and K_1, while E is the initial element of the portion of G between K_0 and K_1. Hence for k sufficiently large, h_k, according to (1), will lie arbitrarily near the portion of G between K_0 and K_1. But the portion of V between K_k and K_{k+1} is, by definition of V_k, congruent to h_k, and accordingly for k sufficiently large lies arbitrarily near the portion of G between K_k and K_{k+1}. Thus V is asymptotic to G in its positive sense. The proof of the theorem is complete.

DEFINITION. The two non-intersecting geodesics of Class A of a given type which form the two boundary curves of the region outside of which no geodesics of Class A of the given type can pass, will be termed the *boundary geodesics* of that type (cf. Theorem 8). The name will be applied even in the special case where the two bounding geodesics are identical.

THEOREM 11. *The boundary geodesics of the type of a periodic* NE *straight line L are themselves periodic.*

Let T be a transformation of the group that carries L into itself. T carries L into itself, and hence every curve of the type of L into another such curve. Further, T, as well as every other transformation of the group, preserves surface distances, and hence carries every geodesic of Class A into a geodesic of Class A. Thus T carries a geodesic of Class A of the type of L into another such geodesic.

To prove the theorem, observe first that the two given boundary geodesics, say G_1 and G_2, divide the unit circle into three non-overlapping regions, say S_1, S_2, and S, of which S lies between G_1 and G_2, S_1 lies between G_1 and the unit circle, and S_2 lies between G_2 and the unit circle. S, in case G_1 is identical

with G_2, will reduce simply to the geodesic $G_1 = G_2$. Now S_1 contains in its interior no geodesic of Class A of the type of L, but is, however, bounded within the unit circle by one such geodesic. These two properties of S_1 must be preserved in the region, say S', into which S_1 is carried by the transformation T. For otherwise, upon applying the inverse of T to S', we could conclude that the two properties attributed to S_1 did not really belong to S_1. Now the only regions of the unit circle which have these properties of S_1, and like S_1 are bounded in part by the unit circle, are the regions S_1 and S_2. Hence the transformation T must carry S_1 either into S_1, or else into S_2. But T cannot carry S_1 into S_2; for that portion of the unit circle that bounds S_1 lies between the two fixed points of T, and under a hyperbolic transformation, such as T, is carried into the same portion of the unit circle. Thus S_1 goes into S_1. Hence G_1, as the boundary of S_1 within the unit circle, goes into G_1. It follows similarly that G_2 is carried into G_2 by T, and the theorem is proved.

THEOREM 12. *If two closed geodesics of Class* A *are of the same type they are also of the same length.*

Since the two given closed geodesics are of the same type they will be represented on the unit circle by an infinite number of pairs of periodic curves, the members of each pair of the same type. Let C and D be the members of one such pair. Let T be a primitive transformation which carries C and D into themselves. Let c be a segment of C of which one end point A is carried into the other end point by T. Let d be a segment of D of which one end point B is carried into the other end point by T. Denote the surface lengths of c and d also by c and d, respectively.

Suppose now that the theorem were false, and that

$$d - c = e > 0.$$

Let h be the surface distance between A and B. Let n be a positive integer such that

$$n > \frac{2h}{e}.$$

Consider now a continuous portion of C, say c', consisting of all of the images of c obtained by applying T^n times to c. Let d' be a similar piece of D obtained by applying T^n times to d'. The end points of d' can be joined by a curve whose surface length is shorter than that of d', by passing from the end point B of d' to the end point A of c' through a surface distance h, passing along c through a surface distance nc, thence returning to the second end point of d' through a surface distance h, in all over a path whose surface

length would be $2h + nc$ which according to the preceding inequalities would be less than nd, or the surface length of d'. This is impossible if D is of Class A. The theorem thus is proved.

18. Geodesics of Class A asymptotic to periodic geodesics of Class A.

THEOREM 13. *A geodesic a of Class* A *which is asymptotic to a periodic geodesic b of Class* A *within the unit circle cannot cross any periodic geodesic of Class* A *of the type of b.*

Suppose that a did cross a periodic geodesic c of Class A of the type of b, at a point P. (See Fig. 5.) Let there be assigned to a, b, and c, positive senses that lead to their common end point on the unit circle. Let m be the length of the closed geodesic which on the original surface corresponds to b.

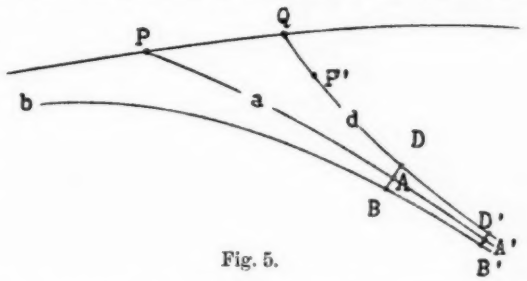

Fig. 5.

According to Theorem 12, the length of the closed geodesic which corresponds to c is also m. On c let there be laid off from P in c's positive sense a segment whose surface length is m; denote the second end point of this segment by Q. The transformation T which carries P into Q, and b and c into themselves, carries a into a geodesic d of Class A that is again asymptotic to b.

Let P' be any point on d, following Q, and neighboring Q. Any curve which contains the corner PQP', formed at Q by c and d, can be shortened in surface length, without disturbing P and P', by a positive constant, say e. Let B and B' be two points on b which are the two end points of a segment of b of surface length m, and which are so remote on b that two geodesics perpendicular to b respectively at B and B' meet the asymptotes a and d in points within a surface distance $e/4$ of B and B' respectively, and cut out of a and d respectively two segments whose surface lengths differ from m by less than $e/4$. Denote the point in which the geodesic perpendicular to b at B meets a by A, and the point at which it meets d by D, and denote the point in which the congruent geodesic perpendicular to b at B' meets a by A', and d by D'.

We will now show that the segment PA on a can be replaced by a curve joining its end points, which has a smaller surface length. To that end we compare the surface length of PA with that of the broken curve $PQDA$. We shall use the letters PA to denote the surface length of PA measured along the geodesic a, and will extend this convention to all other segments to be

considered. Now since the segment PA is congruent to the segment QD' we have

(1) $$PA = QD',$$

or

(2) $$PA = QD + DD'.$$

Now, from the choice of B and B',

(3) $$DD' > BB' - \frac{e}{4}.$$

Adding (2) and (3),

(4) $$PA > QD + BB' - \frac{e}{4}.$$

Now

$$BB' = m = PQ.$$

Hence (4) becomes

(5) $$PA > QD + PQ - \frac{e}{4}.$$

Further, from the choice of B,

(6) $$\frac{e}{4} > DA.$$

Adding (5) und (6)

$$PA > PQ + QD + DA - \frac{e}{2},$$

or

$$PA > PQDA - \frac{e}{2}.$$

Now the broken curve $PQDA$ contains the above mentioned corner at Q, and can be replaced by a curve joining its end points whose surface length will, according to the choice of e, be e less than that of $PQDA$. This new curve will accordingly join the end points of PA, and according to the last inequality be less in surface length than PA contrary to the fact that PA is a segment of Class A. Thus our assumption that a crosses a periodic geodesic of Class A of the type of b is impossible, and the theorem is proved.

One of the hypotheses of Theorem 10 requires that the geodesic V of that theorem be of Class A. However, what is essential in the proof of that

theorem can be carried over to the case where every segment on V from a certain point on is of Class A. In this way we get the following lemma:

LEMMA 10. *A portion of a geodesic which in one sense, say its positive sense, is unending, every segment of which portion is of Class* A, *and which in its positive sense comes and remains at an* NE *distance less than an assignable positive constant from some periodic geodesic b of Class* A *is either asymptotic in its positive sense to a periodic geodesic of the type of b, and of Class* A, *or is itself periodic of the type of b.*

There is a similar difference between Theorem 13 and the following lemma:

LEMMA 11. *A portion of a geodesic which in one sense from a certain point on is of Class* A, *and which is asymptotic in the given sense to a periodic geodesic b of Class* A, *cannot cross any periodic geodesic of Class* A *of the type of b.*

THEOREM 14. *If there exist two different periodic geodesics of Class* A, *namely, b and c, of the same type and between which there are no other periodic geodesics of Class* A, *then if to b and c there be assigned positive senses that lead on b and c to a common end point of b and c on the unit circle, there exist at least two geodesics of Class* A *lying in the region between b and c, of which one geodesic is asymptotic to b in b's positive sense and to c in c's negative sense, while the other geodesic is asymptotic to b in b's negative sense and to c in c's positive sense.*

Let $B_1\ B_2\ B_3 \cdots$ be a set of points which follow one another on b in b's positive sense, and whose surface distances from B_1 become infinite with the subscript n. Let $C_1\ C_2\ C_3 \cdots$ be a set of points on c which follow each other other on c in c's negative sense and whose surface distances from C_1 become infinite with the subscript n. Let h_n be a segment of a geodesic of Class A joining B_n to C_n. Let P_n be a point on h_n whose surface distances from b and c are equal.

Let k be a segment of a geodesic of Class A joining some point of b to some point of c. Let T be a transformation of the fundamental group that carries b and c into themselves. (Cf. Theorem 9.) Under T, k will be carried into a second geodesic segment, say k', also of Class A, and also joining a point of b to one of c. Between k and k', b and c, there will be a region, say S, contained in the region, say S', between b and c, and such that every point of S' is congruent under some multiple of T or its inverse to some point of S.

A proper multiple of T or its inverse will carry P_n into a point P_n' of S. The same transformation will carry h_n into a second geodesic segment of Class A, say h_n', again joining a point of b to a point of c. Let E_n be the element on h_n' at the point P_n'. The elements E_n will have at least one limit element E. Denote the initial point of E by P. Let d be the geodesic defined by E. We

will show that d is asymptotic to b in b's positive sense, and to c in c's negative sense.

The h'_n all become infinite in surface length with n. (Cf. Lemma 3.) Accordingly d will become infinite in surface length, in at least one sense, without passing out of the region between b and c. That portion of d that lies between b and c, and contains the point P, will certainly be of Class A. For any of its finite segments between b and c permit an arbitrarily close approximation, both in position and in surface length, by a properly chosen segment of some one of the segments h'_n, which are themselves of Class A. (Cf. proof of Lemma 9, § 14.) According to Lemma 10, d traced out in that sense in which it is already proved to remain between b and c, becomes asymptotic to some periodic curve of Class A of the type of b and c. According to Lemma 11, d cannot then cross any periodic curve of Class A, of the type of b or c. In particular d cannot cross b or c, and hence remains between b and c. As stated before every segment of d which contains the point P, and remains between b and c, is of Class A. We can conclude therefore that every segment of d is of Class A.

Finally it would be impossible for d, as defined, to be asymptotic to b in b's negative sense, and to c in c's positive sense. For the methods of § 16, applied to any geodesic segment h_n, will show that if h_n be traced from its end point B_n to its end point C_n, its successive intersections with the geodesic segments congruent to k under T will recede from b and progress toward c, in the sense that their surface distances measured from b along the segments congruent to k will increase. Now according to the definition of d, there exists a geodesic segment h'_n, which, starting near the point P on d, will follow along d arbitrarily near d, for an arbitrarily long surface distance, measured on d from P, in either sense. Hence d can only be asymptotic to b in b's positive sense, and to c in c's negative sense.

The existence of a geodesic of Class A asymptotic to b in b's negative sense and to c in c's positive sense is similarly proved.

THEOREM 15. *It is impossible within the unit circle for a geodesic of Class* A *to be asymptotic in both of its senses to the same periodic geodesic of Class* A.

Suppose the theorem were false and that a geodesic b of Class A were asymptotic in both of its senses to a periodic geodesic c of Class A. Upon applying any transformation of the fundamental group which carries c into itself b would be carried into a second geodesic, say b', of Class A also asymptotic in both of its senses to c. Now b and b' would intersect. For if b, for example, lay entirely within the region between b' and c the maximum N E distance of any of b's points from c would be less than that of b'. This maximum, however, should be the same for both b and b' since they are congruent. Thus b would intersect b' contrary to the fact, as stated in Theorem 6,

that no two geodesics which are asymptotic can intersect. Thus the initial assumption of this proof is false, and the theorem is true.

19. **Completion of results on periodic geodesics.** Let L be any periodic N E straight line. Denote the set of all geodesics of Class A and of the type of L by M. The properties of the geodesics of the set M may be summed up and completed as follows:

If there is but one geodesic in the set M it is periodic. If there is more than one geodesic in M there are two non-intersecting geodesics of M, say b_1 and b_2, called boundary geodesics of M, delimiting a region outside of which no geodesics of M can pass. The boundary geodesics are periodic. No periodic geodesics of M intersect. Between b_1 and b_2 there may be a finite or infinite number of periodic geodesics of M or none. If there exists a single point in the region between b_1 and b_2 through which there does not pass a periodic geodesic of M, there is at least one region bounded by two periodic geodesics of M, say p_1 and p_2, between which there are no other periodic geodesics of M. Denote one end point of L on the unit circle by A, and the other end point by B. In the region between p_1 and p_2 there is at least one geodesic of M asymptotic to p_1 as A is approached, and asymptotic to p_2 as B is approached, and at least one geodesic of M asymptotic to p_1 as B is approached and to p_2 as A is approached. Any geodesic of M which is not periodic lies in a region between two periodic geodesics of M, between which there are no periodic geodesics of M. Any such non-periodic geodesic of M is always asymptotic in one sense to one of the periodic geodesics bounding the region in which it lies, and to the other such periodic geodesic in the other sense.

Geodesics of Class A, of the type of N E straight lines that cross L, cross all geodesics of the set M and recede indefinitely from all points of geodesics of M. Any geodesic, say c, of Class A, of the type of any one of the infinite number of N E straight lines asymptotic to L, lies wholly outside the region between b_1 and b_2, and as A is approached becomes asymptotic to that one of the two geodesics, b_1 or b_2, which separates c from the region between b_1 and b_2. A similar statement may be made replacing the point A by the point B.

We conclude with a proof of the following theorem.

THEOREM 16. *The set of all periodic geodesics of Class* A *include among their limit geodesics all of the "boundary" geodesics.* (Cf. § 17.)

Let L be a given N E straight line. According to the lemma of § 8 there exists a transformation of the fundamental group, having its fixed points arbitrarily near the end points of any preassigned arc of the unit circle. The N E straight line passing through the fixed points of a transformation of the group is a periodic straight line. Accordingly it will be possible to choose a set of periodic N E straight lines, say M, which have L as a limit line,

which have no points in common with L either within or on the unit circle, and which all lie on one side of L. Let N be a similar set of periodic N E straight lines all on the other side of L. According to Theorem 11, § 17, the boundary geodesics of the type of any periodic N E straight line are themselves periodic. Accordingly, corresponding to each periodic N E straight line of the set M there exists at least one geodesic of Class A of the same type that is itself periodic; denote the set of such periodic geodesics by M'. Let N' be a similar set of periodic geodesics consisting of one periodic geodesic of Class A of the type of each N E straight line of the set N. Now exactly as was shown in the proof of Theorem 8, § 15. the geodesics of the sets M' and N' will have, the one, a limit geodesic M_0, and the other. a limit geodesic N_0, that are the two boundary geodesics of the geodesics of Class A of the type of L. Thus the theorem is proved.

CORNELL UNIVERSITY,
ITHACA, N. Y.

THE HILBERT INTEGRAL AND MAYER FIELDS FOR THE PROBLEM OF MAYER IN THE CALCULUS OF VARIATIONS*

BY

GILLIE A. LAREW

One studying the problem of Mayer in the calculus of variations and investigating the sufficient conditions is naturally led to inquire whether it may be possible to extend the Hilbert theory to cover this problem, as has been done in the case of the Lagrange problem through the work of A. Mayer and Bolza.† In Kneser's extensive studies of the fields available for the Mayer problem‡ no use is made of an integral analogous to that of Hilbert. In the present paper such an integral is constructed. It is also shown that the conditions for this integral to be independent of the path are equivalent to conditions upon the field of extremals consistent with those found by Kneser and analogous to those characterizing the "Mayer fields" of the Lagrange problem.

1. The Hilbert integral and a Weierstrass theorem

The Mayer problem here considered may be formulated as follows: Among all systems of functions $y_0(x), y_1(x), \ldots, y_n(x)$ which satisfy the $m+1$ differential equations

$$(1) \quad \varphi_\alpha(x, y_0, y_1, \ldots, y_n, y_0', y_1', \ldots, y_n') = 0 \quad (\alpha = 0, 1, \ldots, m;\ m < n)$$

and for which $y_0, \ldots, y_n$ take on fixed values $y_{01}, \ldots, y_{n1}$ at $x = x_1$, while $y_1, \ldots, y_n$ take fixed values $y_{12}, \ldots, y_{n2}$ at $x = x_2$, it is required to determine a system giving $y_0(x_2)$ a minimum. Primes here, and throughout the paper, indicate derivatives with respect to x.

---

* Presented to the Society, December 28, 1920.

† Mayer's three papers *Über den Hilbertschen Unabhängigkeitssatz in der Theorie des Maximums und Minimums der einfachen Integrale,* Berichte über die Verhandlungen der Königlichen Sächsischen Gesellschaft der Wissenschaften zu Leipzig, mathematisch-physikalische Klasse, vol. 55 (1903), pp. 131–145; vol. 57 (1905), pp. 49–67, 313–314. Bolza, these Transactions, vol. 7 (1906), pp. 459–488; *Vorlesungen über Variationsrechnung,* § 78; Rendiconti del Circolo Matematico di Palermo, vol. 31, pp. 257–272.

‡ *Lehrbuch der Variationsrechnung,* Abschnitt VII, §§ 59, 60; Archiv der Mathematik und Physik, vol. 24 (1915), p. 26.

It will be convenient to use indices with the following ranges:

$$i, j = 0, 1, \ldots, n, \qquad r, s, t = 1, 2, \ldots, n,$$
$$\alpha, \beta = 0, 1, \ldots, m.$$

The symbol F will be used to denote the sum

$$F(x, y_0, \ldots, y_n, y_0', \ldots, y_n', \lambda_0, \ldots, \lambda_m) = \sum_\alpha \lambda_\alpha(x)\, \varphi_\alpha(x, y_0, \ldots, y_n, y_0', \ldots, y_n')$$

the λ's being those functions of x sometimes called the Lagrange multipliers. The existence of such functions, not all identically zero on $x_1\, x_2$ and forming with the functions φ_α the Euler-Lagrange equations given below, is a necessary condition for a minimum.* Partial derivatives will be indicated by appropriate subscripts and it will frequently be convenient to introduce the symbols

$$F_i = \partial F/\partial y_i; \qquad G_i = \partial F/\partial y_i'.$$

The Euler-Lagrange equations for the Mayer problem written in these notations are

$$F_i - \frac{d}{dx} G_i = 0; \quad \varphi_\alpha = 0. \tag{2}$$

A solution of these equations will be called an *extremal*.

Consider an n-parameter family of extremals

$$y_i = Y_i(x, a_1, \ldots, a_n), \qquad \lambda_\alpha = \Lambda_\alpha(x, a_1, \ldots, a_n). \tag{3}$$

Let the parameters $a_1, \ldots, a_n$ be restricted to a region $\mathfrak{A}$ and the variable x subject to the condition $\xi_1(a_1, \ldots, a_n) \leqq x \leqq \xi_2(a_1, \ldots, a_n)$, where ξ_1 and ξ_2 are continuous functions of $a_1, \ldots, a_n$ such that $\xi_1(a_1, \ldots, a_n) \leqq \xi_2(a_1, \ldots, a_n)$. Let the extremals, in addition to satisfying equations (2), have the following properties:

(1) The functions y_i, y_i', are of class C'.†

(2) These functions define points for which the φ_α are of class C''.

(3) Along each extremal $G_0 \neq 0$.

(4) The determinant $|\partial y_r/\partial a_s|$ is everywhere different from zero.

* Bolza, *Vorlesungen über Variationsrechnung*, p. 574.

† The word class is used as by Bolza, *Vorlesungen über Variationsrechnung*, pp. 13, 63.

The implicit function theorem establishes the existence of inverse functions

$$a_r = A_r(x, y_1, \ldots, y_n).$$

Define a function

$$\theta_0(x, y_1, \ldots, y_n) = Y_0(x, A_1, \ldots, A_n),$$

so that $\theta_0(x, Y_1, \ldots, Y_n) = Y_0(x, a_1, \ldots, a_n)$. Then

$$y_0 = \theta_0(x, y_1, \ldots, y_n)$$

is the equation of a surface on which lie the extremals (3).

The projections of these extremals then simply cover a region in the space $x, y_1, \ldots, y_n$, which may be called the field $\mathfrak{F}'$. The slope functions and the multiplier functions of $\mathfrak{F}'$ are respectively defined by the sets of equations

$$p_i(x, y_1, \ldots, y_n) = Y_i'(x, A_1, \ldots, A_n),$$
$$\mu_\alpha(x, y_1, \ldots, y_n) = \Lambda_\alpha(x, A_1, \ldots, A_n).$$

The equations $y_0 = \theta_0(x, y_1, \ldots, y_n)$, $y_r = y_r$, $y_0' = \theta_0'(x, y_1, \ldots, y_n)$, $y_r' = p_r(x, y_1, \ldots, y_n)$ define a transformation between the region $\mathfrak{F}'$ and a region in the $(2n+3)$-space $(x, y_0, y_1, \ldots, y_n, y_0', y_1', \ldots, y_n')$ in which, from the hypotheses on the family (3), the functions φ_α are of class C''. The result of substituting the functions θ_0, p_i, μ_α for y_0, y_i', λ_α respectively in the expressions for F, F_i, G_i, and φ_α will be denoted by enclosing the symbols in brackets, as $[F]$.

Definition of Mayer field

The region $\mathfrak{F}'$ will be called a Mayer field, if the integral

$$(4) \qquad I^* = \int_{x_0}^{x} \left\{ -[F/G_0] + \sum_i p_i [G_i/G_0] - \sum_r y_r' [G_r/G_0] \right\} dx$$

is independent of the path in $\mathfrak{F}'$. The integral (4) will be called the Hilbert integral.

The equations

$$(5) \qquad \partial A/\partial y_r - \partial B_r/\partial x = 0; \qquad \partial B_r/\partial y_s - \partial B_s/\partial y_r = 0,$$

where $A = -[F/G_0] + \sum_i p_i [G_i/G_0]$, $B_r = -[G_r/G_0]$, are the conditions that the integral (4) be independent of the path. Since the family (3) satisfies

equations (2), it is true that $[F] = 0$. By the use of this fact and by the addition and subtraction of the expression $\sum_s p_s \, \partial [G_r/G_0]/\partial y_s$ the following set of equations may be derived from the first group of equations (5):

$$
\begin{aligned}
[G_0](\partial A/\partial y_r - \partial B_r/\partial x) = {} & \left\{\left[\frac{d}{dx} G_r\right] - [F_r]\right\} - [G_r/G_0]\left\{\left[\frac{d}{dx} G_0\right] - [F_0]\right\} \\
& - [F_0]\,\{\partial \theta_0/\partial y_r + [G_r/G_0]\} \\
& - [G_0] \sum_s p_s (\partial B_s/\partial y_r - \partial B_r/\partial y_s).
\end{aligned} \tag{6}
$$

Since the arcs are extremals, the first two terms on the right in formula (6) vanish.

THEOREM. *If the derivative F_0 is zero, the necessary and sufficient condition for the region $\mathfrak{F}'$ to be a Mayer field is*

$$\partial [G_s]/\partial y_r = \partial [G_r]/\partial y_s; \tag{7}$$

if F_0 is different from zero, the necessary and sufficient condition for the Mayer field is

$$\partial \theta_0/\partial y_r = -[G_r/G_0]. \tag{8}$$

When F_0 is zero, it appears from the relation (6) that the second group of equations (5) is the necessary and sufficient condition for the Mayer field. Since, for an extremal, the vanishing of F_0 implies that G_0 is a constant, these equations reduce to the form (7) given in the theorem. For the Lagrange problem in n-space regarded as a Mayer problem in $(n+1)$-space, F_0 vanishes. The condition (7) is that with which Bolza begins his study of Mayer fields for the Lagrange problem.*

Suppose F_0 is not identically zero. The relation (6) shows that the equations (8) are necessary conditions for a Mayer field. They are also sufficient, since from them follows the second group of equations (5), and these with (2) and (6) give the first group of (5).

The statement and proof of the Weierstrass theorem which can be established for the extremals of this field should be prefaced by a word as to the form which the Weierstrass E-function assumes in the Mayer problem. The writer has in a former paper† developed as a necessary condition for a strong relative minimum the non-negative character of an E-function defined as follows:

---

*Rendiconti del Circolo Matematico di Palermo, vol. 31, pp. 257-272.

†These Transactions, vol. 20 (1919), p. 1.

$$E = F(x, e_0, \ldots, e_n, y_0', \ldots, y_n', \lambda_0, \ldots, \lambda_m) - F(x, e_0, \ldots, e_n, e_0', \ldots, e_n', \lambda_0, \ldots, \lambda_m) - \sum_i (y_i' - e_i') G_i(x, e_0, \ldots, e_n, e_0', \ldots, e_n', \lambda_0, \ldots, \lambda_m),$$

the curve $y_i = e_i(x)$ being a minimizing arc and the direction $(y_0', \ldots, y_n')$ such that at a point x_3, between x_1 and x_2, $x_3, e_i(x_3)$, and y_i' satisfy the equations (1). This demonstration proceeded on the hypothesis that the derivative G_0 (which does not vanish along the minimizing arc) is negative. For G_0 positive the signs in the E-function should be reversed. In other words, the necessary condition really established was

$$-E/G_0 \geqq 0.$$

THEOREM. *Suppose the arc $E(y_i = e_i(x))$ is an extremal arc and that its projection in the space $x, y_1, \ldots, y_n$ lies in a Mayer field $\mathfrak{F}'$. E passes through $P_1(x_1, y_{i1})$ and its projection in the $(n+1)$-space joins $P_1'(x_1, y_{r1})$ to $P_2'(x_2, y_{r2})$. Let $V(y_i = v_i(x))$ be an arc satisfying the equations $\varphi_\alpha = 0$ and joining P_1 to a point whose projection in the space $x, y_1, \ldots, y_n$ is P_2'. Call the projections of E and V in this space E' and V'. Then*

$$v_0(x_2) = e_0(x_2) + \int_{x_1}^{x_2} \left\{ -\frac{1}{G_0} \right\} E(x, v, p, \mu)\, dx, \tag{9}$$

where E in the Weierstrass E-function.

The definition (4) shows that along the arc E'

$$I_E^* = \int_{x_0}^{x} p_0\, dx = \int_{x_0}^{x} d(e_0(x)).$$

Since $\mathfrak{F}'$ is a Mayer field, the value of the integral I^* along the curve E' is the same as its value along V'. Therefore

$$\begin{aligned} v_0(x_2) &= e_0(x_2) + v_0(x_2) - I_{V_{12}}^* - y_{01} \\ &= e_0(x_2) + \int_{x_1}^{x_2} v_0'(x)\, dx - \int_{x_1}^{x_2} \left\{ -[F/G_0] + \sum_i [G_i/G_0] p_i - \sum_r [G_r/G_0] v_r' \right\} dx \\ &= e_0(x_2) + \int_{x_1}^{x_2} \left\{ [F/G_0] + \sum_i [G_i/G_0] (v_i' - p_i) \right\} dx. \end{aligned}$$

The integrand in the last expression on the right is not affected by writing into it the additional term $-F(x, v_i, v_i', \lambda_a)/G_0(x, v_i, v_i', \lambda_a)$ since this term is identically zero, when the arc V satisfies the equations $\varphi_a = 0$. The formula (9) is thus obtained.

2. The structure of a Mayer field

THEOREM. *Consider an n-parameter family of extremals of the type described in* § 1. *If* $\mathfrak{F}'$ *is a Mayer field in the sense of* § 1, *the sums* $\sum_i [G_i]\, \partial y_i/\partial a_r$ *vanish identically; conversely, if these sums vanish identically, there is at least a neighborhood in which the Hilbert integral is independent of the path.*

The proof of the first part of the theorem is immediate from the condition (8); for

$$\frac{\partial Y_0}{\partial a_r} = \sum_s \frac{\partial \theta_0}{\partial y_s} \frac{\partial Y_s}{\partial a_r} = -\sum_s [G_s/G_0] \frac{\partial Y_s}{\partial a_r}.$$

This is equivalent to

$$\sum_i [G_i] \frac{\partial Y_i}{\partial a_r} = 0.$$

For a proof of the second part of the theorem, consider a set $(x, a_1', \ldots, a_n')$ corresponding to a point of $\mathfrak{F}'$. In a neighborhood of this point there is defined

$$\theta_0(x, y_1, \ldots, y_n) = Y_0(x, A_1, \ldots, A_n),$$

the A_r being the inverse functions of the field. Then, if the sums $\sum_i [G_i] \partial Y_i/\partial a_r$ vanish,

$$\frac{\partial \theta_0}{\partial y_r} = \sum_s \frac{\partial Y_0}{\partial a_s} \frac{\partial A_s}{\partial y_r} = -\sum_{st} [G_t/G_0] \frac{\partial Y_t}{\partial a_s} \frac{\partial A_s}{\partial y_r}.$$

If the identity

$$y_t \equiv Y_t(x, A_1, \ldots, A_n)$$

be differentiated with respect to y_r, it appears that

$$\sum_s \frac{\partial Y_t}{\partial a_s} \frac{\partial A_s}{\partial y_r} = \delta_{tr},$$

the symbol δ_{tr} having the value unity when t is equal to r and zero, when t is different from r. It follows that

$$\frac{\partial \theta_0}{\partial y_r} = -[G_r/G_0].$$

Then at least in a neighborhood of the point corresponding to $a_1', \ldots, a_r'$ the Hilbert integral is independent of the path.

THEOREM. *The sums* $\sum_i G_i \partial Y_i/\partial a_r$ *are independent of* x *along each extremal of the family* (3).

From the hypotheses on Y_i, Y_i', Λ_α it follows that

$$\bar{F}(x, a_1, \ldots, a_n) = F(x, Y_0, \ldots, Y_r, Y_0', \ldots, Y_n', \Lambda_0, \ldots, \Lambda_m) = 0.$$

Suppose all the a's are fixed except a_r and that x is independent of a_r. Then $\partial \bar{F}/\partial a_r = 0$, or

$$\sum_i F_i \partial Y_i/\partial a_r + \sum_i G_i \partial Y_i'/\partial a_r + \sum_\alpha \varphi_\alpha \partial \Lambda_\alpha/\partial a_r = 0.$$

Since the equations $\varphi_\alpha = 0$ are satisfied, this can be written

$$\sum_i \left(F_i - \frac{d}{dx} G_i\right) \frac{\partial Y_i}{\partial a_r} + \frac{d}{dx} \sum_i G_i \frac{\partial Y_i}{\partial a_r} = 0.$$

Since equations (2) are satisfied by these extremals, the statement of the theorem is established.

COROLLARY. *If the expression under the integral sign in* (4) *is an exact differential for a special value of* x, *say* $x = x_0$, *it is so for any value of* x.

It is thus clear that, in order for the expression of (4) to be an exact differential for the entire field, it is necessary and sufficient that the conditions $\sum_i [G_i] \partial Y_i/\partial a_r = 0$ shall be satisfied at the intersection of the field with the "hypersurface" $x = x_0$.

RANDOLPH-MACON WOMAN'S COLLEGE,
LYNCHBURG, VA.

NORMAL CONGRUENCES AND QUADRUPLY INFINITE SYSTEMS OF CURVES IN SPACE*

BY

JESSE DOUGLAS

INTRODUCTION

The present paper sets into relation with one another the following two geometric configurations:

First, the quadruply infinite system of curves in space, as defined by the set of differential equations†

$$(1)\qquad \begin{aligned} y'' &= F(x, y, z, y', z'), \\ z'' &= G(x, y, z, y', z'). \end{aligned}$$

Second, the normal congruence — a congruence of curves being any system of ∞^2 curves so distributed that one and only one passes through each point of some region of space, and a normal congruence one relative to which a system of ∞^1 orthogonal surfaces can be constructed. The congruence defined by the differential equations

$$\frac{dx}{A(x, y, z)} = \frac{dy}{B(x, y, z)} = \frac{dz}{C(x, y, z)}$$

is normal when and only when the Pfaffian equation

$$A\,dx + B\,dy + C\,dz = 0$$

is exact—thus a congruence is in general not normal.

The simplest example of a system of ∞^4 curves in space is presented by the straight lines, one of the important occurrences of which is as the extremals of the calculus of variations problem $\int ds =$ minimum.

---

* Presented to the Society, February 25, 1922 (Part I) and April 28, 1923 (Part II).

† The functions F and G are assumed to be uniform and analytic in a certain region of the x, y, z, y', z' continuum, within which region we always remain.

Consider the straight lines orthogonal to an arbitrary surface Σ. They are orthogonal not only to Σ but to its ∞^1 parallel surfaces, obtained by laying off along the lines from their points of normality to Σ constant lengths, that is, constant $\int ds$.

This property of the extremals of $\int ds$ = minimum generalizes to the extremals of any variation problem of the form

$$\int F(x, y, z)\, ds = \text{minimum}. \tag{2}$$

This is because for integrals of that form (and for no others) the relation of transversality in the sense of the calculus of variations reduces to orthogonality. Therefore, according to the classic theory, if we select from the ∞^4 extremals the ∞^2 that meet an arbitrary surface Σ orthogonally, and lay off along each an arc starting from Σ so that $\int F(x, y, z)\, ds$ over the arc has a fixed value, the locus of the end points is a surface orthogonal to the ∞^2 extremals. By variation of the fixed value, ∞^1 such orthogonal surfaces result.

A system of ∞^4 curves in space identifiable with the totality of extremals of a problem of the form (2) is known in the literature as a *natural family*,* the name referring to the many physical interpretations. The Euler-Lagrange equations of (2) are

$$\begin{aligned} y'' &= (L_y - y' L_x)(1 + y'^2 + z'^2), \\ z'' &= (L_z - z' L_x)(1 + y'^2 + z'^2), \end{aligned} \tag{3}$$

where $L = \log F$, and is thus any function of x, y, z. The form of these differential equations may be taken as the criterion for a natural family.

The above stated property of the extremals of $\int F(x, y, z)\, ds$ may be formulated as follows: *in a natural family the ∞^2 curves orthogonal to an arbitrary surface form a normal congruence.* This theorem is due to Thomson and Tait, who stated it in a dynamical form.†

In a paper of 1910,‡ E. Kasner dealt with a number of geometrical problems arising from the theorem of Thomson and Tait, proving, in particular, that

* After E. Kasner. See his *Differential-geometric aspects of dynamics, Princeton Colloquium Lectures*, 1912, § 28.

† *Natural Philosophy* (edition of 1879), Part I, § 332. In its optical form, the theorem is known as that of Malus; see Hadamard, *Leçons sur le Calcul des Variations*, p. 150.

‡ *The theorem of Thomson and Tait and natural families of trajectories*, these Transactions, vol. 11 (1910), pp. 121–140.

the property of the theorem is *characteristic* of natural families. The writer's attention was directed to this field of investigation by Professor Kasner, the problems of the present paper being suggested by the paper just cited. Our method of analysis is distinguished by the systematic use of equations of variation.

Our paper falls into two parts. The first determines all the possibilities as to the number of normal congruences—in the sense of the number of parameters or of arbitrary functions involved—that may be contained in a quadruply infinite system of curves in space. The second deals with special base manifolds Σ, solving the four problems of determining all families of ∞^4 space curves, such that a normal congruence is formed by (1) the curves of the family passing through an arbitrary point, (2) the curves orthogonal to an arbitrary plane, (3) the curves orthogonal to an arbitrary sphere, (4) the curves orthogonal to an arbitrary straight line.

Given a family of ∞^4 curves in space, every surface Σ determines a congruence of curves within the family, consisting namely of the ∞^2 curves of the family orthogonal to Σ. In the case of a natural family, all these congruence have the normal character. At the other extreme is the general family of ∞^4 curves which contains no normal congruences whatever. The investigations that follow bring to light a number of intermediate types, resulting in the following complete classification.

1. The natural families, which are such that the congruence determined by an arbitrary surface Σ is normal.

2. The quasi-natural families. These are such that all and only the surfaces Σ which obey a certain linear partial differential equation of the second order

$$\mathfrak{B}_1 r + \mathfrak{B}_2 s + \mathfrak{B}_3 t + \mathfrak{B}_4 = 0$$

give rise to normal congruences. The coefficients are functions of x, y, z, p, q whose form depends on the particular family.

3. A type in which, if there is given arbitrarily a space curve together with a surface element belonging to the curve, there exists a unique surface Σ containing curve and element and determining a normal congruence.

4. A type containing ∞^3 normal congruences, so distributed that each curve of the family belongs to ∞^1 of them.

5. A type containing ∞^2 normal congruences, so distributed that each curve of the family belongs to at least one of them.

6. The normal congruences may be infinite in number, but confined to a triply infinite sub-family. In a triply infinite curve family in space the directions of the curves through an arbitrary point form a cone. It follows that in order that a surface Σ be met by ∞^2 curves of the family orthogonally,

it must satisfy a certain partial differential equation of the first order. Either all these surfaces give normal congruences, or else ∞^2, or ∞^1, or none of them; so that, since each normal congruence absorbs ∞^1 surfaces Σ, there are, if more than a finite number, either ∞^∞ or ∞^1 normal congruences in the family.

7. A finite number or no normal congruences.

The base manifold Σ may reduce to a curve or to a point. In the paper cited, Professor Kasner proved it sufficient to characterize a system of ∞^4 curves as of the natural type that the curves belonging to it which meet an arbitrary base curve orthogonally form a normal congruence. In fact, he proved that it suffices to demand this property of only the ∞^6 circles of space.

Relative to the case where the base manifold is a point, Professor Kasner showed by means of the example

$$y'' = y'^2, \qquad z'' = 0$$

that non-natural families exist in which the curves through an arbitrary point of space form a normal congruence*, and formulated but left unsolved the problem of determining all such families, with the remark that the direct attack on the problem by the methods of his paper would lead to great difficulties.

The second part of the present paper solves this problem together with the analogous ones relative to the plane, sphere, and straight line. The results are as follows.

1. There are three types with the normality property with respect to points: the natural, the quasi-natural, and a third type defined by a certain system of four partial differential equations in F and G (the equations (43), § 8).

2. For the plane as base manifold, the analysis yields five possibilities, of three of which we have established the effective existence. Two of these are the natural and quasi-natural types.

3. The congruence determined by an arbitrary sphere is normal when and only when the family is natural or quasi-natural. This result affords the neatest characterization of a quasi-natural family: one in which the curves orthogonal to any sphere form a normal congruence, but where the same is not true of an arbitrary surface. It is sufficient to demand the normality property of arbitrary points and planes.

4. If the curves orthogonal to an arbitrary straight line form a normal congruence, the family is necessarily natural. It is sufficient to demand this

* The example cited is arrived at analytically. A more general example, based on geometrical considerations, may be obtained by allowing the moulure surface M mentioned in § 6 of the present paper to reduce, as it can, to an arbitrary point.

property of any triple complex of straight lines (three lines in each pencil), or of a triple complex of curves of any kind. In fact, if the two complexes of straight lines parallel to two of the coördinate planes, respectively, give rise to normal congruences, the family must be natural.

In this way, $2\,\infty^3$ tests as to the possession of the normality property are sufficient to decide as to the natural or non-natural character of a given system of ∞^4 curves. An interesting problem is that of finding a minimum number of tests necessary and sufficient for this purpose,—it is certain that this minimum number is either $2\,\infty^3$ or $1\,\infty^3$, but the calculations required to establish a proof for either case seem too complicated. We can be sure, however, that if $1\,\infty^3$ base curves do suffice, they cannot be any whatever, for the family

$$y'' = \frac{y'(1+y'^2+z'^2)}{x+zz'}, \qquad z'' = 0$$

—consisting of the ∞^4 circles whose axes intersect the axis of y, assumed vertical, at right angles—which family is not natural (or even quasi-natural), is such that the $1\,\infty^3$ horizontal straight lines give rise to normal congruences.

There remain these two possibilities as to which we have not been able to decide: (1) perhaps an *arbitrary* system of $2\,\infty^3$ curves giving rise to normal congruences will characterize a family as natural; (2) perhaps $1\,\infty^3$ curves are sufficient, provided they are not of a special type.

Part I

Classification of quadruply infinite systems of curves in space with respect to the normal congruences contained within them

1. Equations of variation

Our analysis being based on equations of variation, we develop the necessary formulas immediately in order to avoid later digression.

In the four-parameter family of curves defined by the equations (1) let

$$y = \varphi(x; \alpha, \beta),$$

$$z = \psi(x; \alpha, \beta)$$

represent any two-parameter sub-family. Then we must have identically in x, α, β

$$\varphi'' = F(x, \varphi, \psi, \varphi', \psi'),$$

$$\psi'' = G(x, \varphi, \psi, \varphi', \psi'),$$

where the accents denote partial differentiation with respect to x. Differentiating these identities with respect to α, and taking account of the commutativity of the differentiations with respect to x and α, we get

$$(\varphi_\alpha)'' = F_y(\varphi_\alpha) + F_z(\psi_\alpha) + F_{y'}(\varphi_\alpha)' + F_{z'}(\psi_\alpha)',$$
$$(\psi_\alpha)'' = G_y(\varphi_\alpha) + G_z(\psi_\alpha) + G_{y'}(\varphi_\alpha)' + G_{z'}(\psi_\alpha)'.$$

Thus the parametric derivatives φ_α, ψ_α obey a system of linear differential equations. Evidently, the same equations are obeyed by φ_β, ψ_β.

2. Conditions for a normal congruence

Any congruence of curves is definable by differential equations of the form

$$\frac{dx}{A(x, y, z)} = \frac{dy}{B(x, y, z)} = \frac{dz}{C(x, y, z)}. \tag{4}$$

A system of ∞^1 surfaces orthogonal to the congruence exists when and only when the corresponding Pfaffian equation

$$A\,dx + B\,dy + C\,dz = 0$$

is exact; for this the necessary and sufficient condition is

$$A(C_y - B_z) + B(A_z - C_x) + C(B_x - A_y) = 0.$$

The equations (4), assigning to each point of space a determinate direction, define a field of ∞^3 lineal elements wherefrom the congruence of curves arises by integration. The most general way of representing such an element field is by writing the five coördinates x, y, z, y', z' of a lineal element as functions of three parameters t, u, v. The condition that an element field so defined correspond to a normal congruence is

$$(y'xy) - (z'zx) + y'(z'yz) - z'(y'yz) = 0, \tag{5}$$

where the parentheses denote jacobians with respect to t, u, v.

3. The congruence determined by an arbitrary base surface

Consider the general family $\mathfrak{F}$ of ∞^4 curves,

$$y'' = F(x, y, z, y', z'),$$
$$z'' = G(x, y, z, y', z').$$

Let Σ designate an arbitrary surface, whose equation we write in the form

$$x = -\varphi(y, z).$$

For the derivatives of φ we shall employ the Monge notation

$$p = \varphi_y, \quad q = \varphi_z; \quad r = \varphi_{yy}, \quad s = \varphi_{yz}, \quad t = \varphi_{zz}.$$

We have for the direction $1 : y' : z'$ normal to Σ at the point (x, y, z) of Σ

$$y' = p, \quad z' = q.$$

The point and direction $x, y, z, y' = p, z' = q$ determine in the family $\mathfrak{F}$ a unique curve C, whose equations can be written explicitly in the form of power series; these begin

$$(6) \qquad \begin{aligned} Y &= y + p(X - x) + \frac{1}{2} F(X - x)^2 + \cdots, \\ Z &= z + q(X - x) + \frac{1}{2} G(X - x)^2 + \cdots. \end{aligned}$$

Here X, Y, Z are the coördinates of a point current along C, and the substitutions

$$x = -\varphi(y, z), \quad p = \varphi_y(y, z), \quad q = \varphi_z(y, z)$$

are to be made throughout, so that the symbol F, for instance, is an abbreviation for

$$F[-\varphi(y, z), y, z, \varphi_y(y, z), \varphi_z(y, z)].$$

In this way, the second members of (6) are functions of X, y, z:

$$(7) \qquad \begin{aligned} Y &= Y(X; y, z), \\ Z &= Z(X; y, z). \end{aligned}$$

The parameters y, z may be regarded as Gaussian coördinates on Σ, and the equations (7) define, then, for fixed y, z, the curve of $\mathfrak{F}$ which is orthogonal to Σ at the point of coördinates y, z—and when y and z are variable, the congruence of curves in $\mathfrak{F}$ determined by orthogonality to Σ.

Adjoining to (7) the equations obtained by differentiation with respect to X, and also the identity $X = X$, we have in the five equations

$$X = X,\ Y = Y(X,y,z),\ Z = Z(X,y,z),\ Y' = Y'(X,y,z),\ Z' = Z'(X,y,z),$$

a representation of the field of lineal elements determined by the congruence in terms of X, y, z as parameters. The condition (5) that the congruence be normal here takes the form

$$\begin{aligned}(Y'G - Z'F)(YZ) + (1+Z'^2)(YY') - Y'Z'(YZ') - Y'Z'(ZY') \\ -(1+Y'^2)(Z'Z) = 0,\end{aligned} \tag{8}$$

where the parentheses such as (YZ) represent determinants in the matrix

$$\begin{vmatrix} Y_y & Z_y & Y'_y & Z'_y \\ Y_z & Z_z & Y'_z & Z'_z \end{vmatrix}.$$

We introduce the notation

$$\omega_1 = (YZ), \quad \omega_2 = (YY'), \quad \omega_3 = (YZ'),$$
$$\omega_4 = (ZY'), \quad \omega_5 = (Z'Z), \quad \omega_6 = (Y'Z'),$$

by means of which (8) abbreviates to

$$(Y'G - Z'F)\omega_1 + (1+Z'^2)\omega_2 - Y'Z'\omega_3 - Y'Z'\omega_4 - (1+Y'^2)\omega_5 = 0. \tag{9}$$

In all that follows we think of y and z as having fixed though arbitrary values, determining thus a single but arbitrary curve C of the congruence orthogonal to Σ. *All the functions of the analysis become in this way functions of the one variable X, the abscissa along C.*

4. The linear differential equations of the ω's

The next step is to show how the six ω's may be uniquely defined by a certain system of linear differential equations together with certain initial conditions.

We have under consideration a congruence of curves

$$Y = Y(X; y, z),$$

$$Z = Z(X; y, z)$$

selected from the totality of curves obeying the differential equations

$$Y'' = F(X, Y, Z, Y', Z'),$$

$$Z'' = G(X, Y, Z, Y', Z').$$

It follows by § 1 that the parametric derivatives Y_y, Z_y obey the linear differential equations

$$(10_1) \qquad \begin{aligned} (Y_y)'' &= F_y(Y_y) + F_z(Z_y) + F_{y'}(Y_y)' + F_{z'}(Z_y)', \\ (Z_y)'' &= G_y(Y_y) + G_z(Z_y) + G_{y'}(Y_y)' + G_{z'}(Z_y)',* \end{aligned}$$

and that Y_z, Z_z obey the same differential equations:

$$(10_2) \qquad \begin{aligned} (Y_z)'' &= F_y(Y_z) + F_z(Z_z) + F_{y'}(Y_z)' + F_{z'}(Z_z)', \\ (Z_z)'' &= G_y(Y_z) + G_z(Z_z) + G_{y'}(Y_z)' + G_{z'}(Z_z)'. \end{aligned}$$

From (10_1) and (10_2) the deduction is direct that the six ω's obey a system of linear equations of the first order. Consider, for example,

$$\omega_2 = \begin{vmatrix} Y_y & Y_y' \\ Y_z & Y_z' \end{vmatrix},$$

and differentiate it with respect to X:

$$\omega_2' = \begin{vmatrix} Y_y & Y_y'' \\ Y_z & Y_z'' \end{vmatrix}.$$

* For convenience of writing, we use small letters as subscripts instead of capitals in denoting the partial derivatives of F and G.

When in the last determinant we substitute for Y_y'', Z_z'' by the equations (10_1), (10_2), we find

$$\omega_2' = F_z \begin{vmatrix} Y_y & Z_y \\ Y_z & Z_z \end{vmatrix} + F_{y'} \begin{vmatrix} Y_y & Y_y' \\ Y_z & Y_z' \end{vmatrix} + F_{z'} \begin{vmatrix} Y_y & Z_y' \\ Y_z & Z_z' \end{vmatrix},$$

that is,

$$\omega_2' = F_z \omega_1 + F_{y'} \omega_2 + F_{z'} \omega_3.$$

In similar manner the derivative of each ω is expressible as a linear combination of ω's, giving the system of equations

$$\begin{aligned}
\omega_1' &= \omega_3 - \omega_4, \\
\omega_2' &= F_z \omega_1 + F_{y'} \omega_2 + F_{z'} \omega_3, \\
\omega_3' &= G_z \omega_1 + G_{y'} \omega_2 + G_{z'} \omega_3 + \omega_6, \\
\omega_4' &= -F_y \omega_1 + F_{y'} \omega_4 - F_{z'} \omega_5 - \omega_6, \\
\omega_5' &= G_y \omega_1 - G_{y'} \omega_4 + G_{z'} \omega_5, \\
\omega_6' &= -G_y \omega_2 + F_y \omega_3 - G_z \omega_4 - F_z \omega_5 + (F_{y'} + G_{z'}) \omega_6.
\end{aligned} \tag{11}$$

The ω's, as solutions of a system of differential equations of the first order, will be completely determined if we know their values for an initial value of X. To find these, we use the power series expansions (6) of the equations of C,

$$Y = y + p(X-x) + \frac{1}{2} F(X-x)^2 + \cdots,$$

$$Z = z + q(X-x) + \frac{1}{2} G(X-x)^2 + \cdots.$$

We recall that x and the coefficients $p, q, F, G, \ldots$ are functions of y and z, coördinates on the surface

$$\Sigma\colon\ x = -\varphi(y, z),$$

and that $p, q; r, s, t$ are used to represent the partial derivatives of the first and second orders of the function φ. With this in mind, we find, on differentiating the equations (7) with respect to y,

$$Y_y = \overline{1+p^2} + \overline{r+pF}\,(X-x) + \cdots,$$

$$Z_y = pq + \overline{s+pG}\,(X-x) + \cdots.$$

Therefore, for $X = x$—marking the point of intersection of C with Σ—

$$Y_y = 1+p^2, \quad Z_y = pq, \quad Y'_y = r+pF, \quad Z'_y = s+pG.$$

Similarly, for $X = x$,

$$Y_z = pq, \quad Z_z = 1+q^2, \quad Z'_y = s+qF, \quad Z'_z = t+qG.$$

Hence the values of the ω's for $X = x$ are

$$(12)\qquad \begin{aligned}
\alpha_1 &= 1+p^2+q^2,\\
\alpha_2 &= qF - pqr + (1+p^2)s,\\
\alpha_3 &= qG - pqs + (1+p^2)t,\\
\alpha_4 &= -pF - (1+q^2)r + pqs,\\
\alpha_5 &= pG + (1+q^2)s - pqt,\\
\alpha_6 &= qGr - (qF+pG)s + pFt + (rt-s^2).
\end{aligned}$$

5. The infinite system of Monge-Ampère equations

The essential elements in the present analysis are three in number:

(i) the system of linear differential equations (11);

(ii) the initial values α of the ω's;

(iii) the linear relation (9).

The necessary and sufficient condition that the congruence of curves selected from the quadruply infinite family $\mathfrak{F}$ by the condition of orthogonality to Σ be a normal congruence is that the six functions ω that obey the differential equations (11), *and take for $X = x$ the values α, shall satisfy the linear relation* (9).

In order to obtain conciseness and symmetry of notation, we condense the equations (11) to

$$(13)\qquad \begin{aligned}
\omega'_1 &= p_{11}\omega_1 + \cdots + p_{16}\omega_6,\\
&\cdots\cdots\cdots\\
\omega'_6 &= p_{61}\omega_1 + \cdots + p_{66}\omega_6,
\end{aligned}$$

and rewrite the linear condition (9) as

$$(14)\qquad A_1\omega_1 + \cdots + A_6\omega_6 = 0,$$

using, that is, the abbreviations

$$(15)\qquad \begin{aligned}
A_1 &= Y'G - Z'F, \quad A_2 = 1+Z'^2, \quad A_3 = -Y'Z',\\
A_4 &= -Y'Z', \quad A_5 = -(1+Y'^2), \quad A_6 = 0.
\end{aligned}$$

Our requirement is that the relation (14) subsist identically in X when the ω's have the values defined by the differential equations (13) together with the initial values (12). A necessary and sufficient condition that (14) subsist identically in X is that the function of X

$$\Omega \equiv A_1 \omega_1 + \cdots + A_6 \omega_6 ,$$

together with its derivatives of all orders, vanish for $X = x$.

All these derivatives are, like Ω itself, linear expressions in the ω's. In fact, when in the first derivative of Ω,

$$\Omega' \equiv A_1 \omega_1' + \cdots + A_6 \omega_6' + A_1' \omega_1 + \cdots + A_6' \omega_6 ,$$

we substitute for the ω''s by (11), we find

$$\Omega' \equiv B_1 \omega_1 + \cdots + B_6 \omega_6 ,$$

where

$$(16_1) \qquad B_i = A_i' + \sum_{j=1}^{j=6} p_{ji} A_j .$$

By repeated differentiation and use of (11) we obtain the infinite sequence

$$\begin{aligned} \Omega &\equiv A_1 \omega_1 + \cdots + A_6 \omega_6 , \\ \Omega' &\equiv B_1 \omega_1 + \cdots + B_6 \omega_6 , \\ \Omega'' &\equiv C_1 \omega_1 + \cdots + C_6 \omega_6 , \\ \Omega''' &\equiv D_1 \omega_1 + \cdots + D_6 \omega_6 , \\ &\cdots\cdots\cdots\cdots , \end{aligned} \qquad (17)$$

where the coefficients in each linear expression are derived from those in the preceding by the same formulas that give the B's in terms of the A's:

$$(16_2) \qquad C_i = B_i' + \sum_{j=1}^{j=6} p_{ji} B_j ,$$

$$(16_3) \qquad D_i = C_i' + \sum_{j=1}^{j=6} p_{ji} C_j ,$$

$$\cdots\cdots\cdots .$$

In demanding that

$$\Omega,\ \Omega',\ \Omega'',\ \ldots$$

all vanish for $X = x$, we get the infinite system of equations

$$\begin{aligned} A_1\alpha_1 + \cdots + A_6\alpha_6 &= 0, \\ B_1\alpha_1 + \cdots + B_6\alpha_6 &= 0, \\ C_1\alpha_1 + \cdots + C_6\alpha_6 &= 0, \\ \cdots\cdots\cdots\cdots & \end{aligned} \tag{18}$$

where the $A, B, C, \ldots$ of these equations are *the values for* $X = x$ of the same letters in (17).

The $A, B, C, \ldots$ of (17) are functions of X, Y, Z, Y', Z', these being relative to a point moving along the curve C. The value $X = x$ corresponds to the intersection of C with Σ, where we have

$$X = x, \quad Y = y, \quad Z = z, \quad Y' = p, \quad Z' = q. \tag{19}$$

Thus the $A, B, C, \ldots$ of (18) are derived from the $A, B, C, \ldots$ of (17) by the substitution (19)—they are functions of x, y, z, p, q, coördinates and first partial derivatives relative to Σ. In (18) and hereafter we shall assume the substitution (19) to be sufficiently well indicated by the context, without employing for it any special notation.

The accent in the formulas (16) denotes total differentiation with respect to X, so that applied to any function of X, Y, Z, Y', Z', it is equivalent to the operator

$$\frac{\partial}{\partial X} + Y'\frac{\partial}{\partial Y} + Z'\frac{\partial}{\partial Z} + Y''\frac{\partial}{\partial Y'} + Z''\frac{\partial}{\partial Z'},$$

or, after the use of the differential equations (1) of C and the substitution (19), to the operator

$$\frac{\partial}{\partial x} + p\frac{\partial}{\partial y} + q\frac{\partial}{\partial z} + F\frac{\partial}{\partial p} + G\frac{\partial}{\partial q}.$$

Rewritten in accordance with (19), the values (15) of the letters A become

$$\begin{aligned} A_1 &= pG - qF, & A_2 &= 1 + q^2, & A_3 &= -pq, \\ A_4 &= -pq, & A_5 &= -(1 + p^2), & A_6 &= 0. \end{aligned} \tag{20}$$

The recurrence formulas (16) are, in expanded form,

$$
\begin{aligned}
B_1 &= A_1' \quad + F_z A_2 + G_z A_3 - F_y A_4 + G_y A_5, \\
B_2 &= A_2' \quad + F_p A_2 + G_p A_3 \quad - G_y A_6, \\
B_3 &= A_3' + A_1 + F_q A_2 + G_q A_3 \quad + F_y A_6, \\
B_4 &= A_4' - A_1 \quad + F_p A_4 - G_p A_5 - G_z A_6, \\
B_5 &= A_5' \quad - F_q A_4 + G_q A_5 - F_z A_6, \\
B_6 &= A_6' \quad + A_3 \quad - A_4.
\end{aligned} \tag{21}
$$

These furnish a definite mechanism whereby the values of the symbols A, B, C, ... may be successively determined. These symbols are seen to be expressions of higher and higher order in the partial derivatives of F and G.

We shall calculate, in addition to the A's, the explicit values of the B's only, which are

$$
\begin{aligned}
B_1 &= pG_x - qF_x - G_y + F_z + p(FG_p + GG_q) - q(FF_p + GF_q), \\
B_2 &= 2qG + (1+q^2)F_p - pqG_p, \\
B_3 &= -2qF + (1+q^2)F_q - pqG_q, \\
B_4 &= -2pG - pqF_p + (1+p^2)G_p, \\
B_5 &= -2pF + pqF_q - (1+p^2)G_q, \\
B_6 &= 0.
\end{aligned} \tag{22}
$$

On substituting the values (12) of the α's in (18), we obtain the following infinite system of partial differential equations of the Monge-Ampère type:

$$
\begin{aligned}
&\mathfrak{A}_0(rt - s^2) + \mathfrak{A}_1 r + \mathfrak{A}_2 s + \mathfrak{A}_3 t + \mathfrak{A}_4 = 0, \\
&\mathfrak{B}_0(rt - s^2) + \mathfrak{B}_1 r + \mathfrak{B}_2 s + \mathfrak{B}_3 t + \mathfrak{B}_4 = 0, \\
&\mathfrak{C}_0(rt - s^2) + \mathfrak{C}_1 r + \mathfrak{C}_2 s + \mathfrak{C}_3 t + \mathfrak{C}_4 = 0, \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned} \tag{23}
$$

where

$$\mathfrak{A}_0 = A_6,$$

$$\mathfrak{A}_1 = -pqA_2-(1+q^2)A_4+qGA_6,$$

$$\mathfrak{A}_2 = (1+p^2)A_2-pqA_3+pqA_4+(1+q^2)A_5-(qF+pG)A_6, \tag{24}$$

$$\mathfrak{A}_3 = (1+p^2)A_3-pqA_5+pFA_6,$$

$$\mathfrak{A}_4 = (1+p^2+q^2)A_1+qFA_2+qGA_3-pFA_4+pGA_5,$$

and the values of $\mathfrak{B}, \mathfrak{C}, \ldots$ are derived respectively from those of $B, C, \ldots$ by the same formulas.

A surface gives rise to a normal congruence when and only when it satisfies the infinite system (23).

The important cases are those where there exists at least one solution of (23) corresponding to arbitrary initial values of x, y, z, p, q, that is, where the system (23) is *completely integrable*. It is easy to see that otherwise any normal congruences present in our four-fold infinitude of curves are either only finite in number, or else confined to a triply infinite sub-family. The situation in the latter case is as described in the sixth of the first group of results collected in the introduction.

6. Classification. The infinite matrix. Quasi-natural families

The classification of quadruply infinite families of curves with respect to the normal congruences contained within them depends on the rank of the infinite matrix

$$\begin{vmatrix} A_1 & \cdots & A_6 \\ B_1 & \cdots & B_6 \\ C_1 & \cdots & C_6 \\ \cdots & \cdots & \cdots \end{vmatrix},$$

together with the complete integrability of the system formed by a certain finite number of the equations (23).

For instance, suppose that the D row of the matrix is a linear combination of the preceding rows:

$$D_i = \varrho A_i+\sigma B_i+\tau C_i.$$

Then it is easily shown by the recurrence formulas (16) that every subsequent row must be a linear combination of its preceding rows; in other words, for the matrix to be of rank three it is necessary and sufficient that

$$|ABCD| = 0, \qquad |ABC| \neq 0,$$

where in writing a matrix equal to zero we mean that every determinant resulting from it by the suppression of columns is equal to zero.

The subsistence of the first three equations of (23) will then imply that of all the succeeding ones. The problem of the complete integrability of (23) is thus referred to that of its first three equations, a matter dealt with in the standard theory.*

It is to be observed that

$$(23_1) \qquad \mathfrak{A}_0 (r t - s^2) + \mathfrak{A}_1 r + \mathfrak{A}_2 s + \mathfrak{A}_3 t + \mathfrak{A}_4 = 0$$

is in reality not present, for by (24) and (20) it disappears identically. This results from the fact that the congruence is constructed orthogonal to Σ. In the case of rank r, therefore, all and only the surfaces obeying the $r-1$ equations (23_2) to (23_r) give normal congruences.

In particular, when

$$(25) \qquad |AB| = 0$$

$(r = 1)$, an arbitrary surface will produce a normal congruence, and conversely. The condition (25) furnishes by (20), (22) a system of partial differential equations for F and G, whose solution gives for those functions the forms (3), characteristic of natural families. We thus have the theorem of Thomson and Tait together with the converse theorem of Kasner stated in the introduction.

The most interesting new case is the next one, of rank two,

$$(26) \qquad |ABC| = 0, \qquad |AB| \neq 0.$$

The families defined by these conditions we term *quasi-natural.*

Their characteristic property is that every surface satisfying the equation (23_2) (which actually is linear, since in fact $\mathfrak{B}_0 = 0$) determines a normal

* E. Goursat, *Leçons sur l'Intégration des Équations aux Dérivées partielles du second Ordre*, tome II, chapter VI.

congruence, while the same is not true of an arbitrary surface. Since the subsistence of (23_2) means that the congruence orthogonal to the surface Σ has an orthogonal surface consecutive to Σ, a quasi-natural family may be described as a non-natural family in which every congruence admitting two consecutive orthogonal surfaces is normal.

It is evidently significant to inquire as to consistency of the conditions (26). The following example, establishing the existence of quasi-natural families, may be of interest.

Suppose that a family of ∞^4 curves in space has these two properties:

(1) its curves are vertical plane curves;

(2) it possesses cylindrical symmetry with respect to every vertical line; that is, every curve of the family remains in the family when rotated about any vertical axis.

For instance, the ∞^4 vertical circles of a given radius.

By solving a simple functional equation, it can be shown that such families are characterized by differential equations of the form

$$\begin{aligned} y'' &= y'^2 f\left(\frac{1+z'^2}{y'^2}, y\right), \\ z'' &= 0 \end{aligned} \tag{27}$$

(the axis of y being assumed vertical). Provided that

$$f \neq \left(\frac{1+z'^2}{y'^2}+1\right) f_1(y), \tag{28}$$

the family is not natural.

Consider an arbitrary vertical plane H, together with an arbitrary curve Γ in the plane. H contains ∞^2 curves of the family, of which ∞^1 meet Γ at right angles. Imagine H to be rolled on a vertical cylinder of arbitrary cross section. Then Γ generates a moulure surface M. The rolling is equivalent to a succession of infinitesimal rotations about the vertical elements of the

* The sub-case where f does not involve y occurs in Professor Kasner's paper (p. 136), being arrived at from an entirely different point of view. The statement on page 136 to the effect that not all curve systems of the form $y''=y'^2 f((1+z'^2)/y'^2)$, $z''=0$ have the property that an arbitrary point taken as base determines a normal congruence, is an error, as follows from the fact that the moulure surface M which we describe below may be allowed to degenerate to an arbitrary point. The Pfaffian on page 137, loc. cit., has an integrating factor.

cylinder. Evidently, therefore, the ∞^1 curves of the family meeting Γ orthogonally in its plane generate ∞^2 curves which belong to the family and are orthogonal to M.

Now, the ∞^1 curves in Π that are orthogonal to Γ admit ∞^1 orthogonal trajectories. In the rolling of Π these generate ∞^1 moulure surfaces at right angles to the ∞^2 curves orthogonal to M; in other words, these ∞^2 curves are a normal congruence.

It is only necessary to remark that the vertical moulure surfaces of space are defined by the linear partial differential equation of the second order

$$(1+q^2)s - pqt = 0$$

to complete the characterization of the families defined by (27), (28) as quasi-natural.

We omit the details of the study of the higher ranks of the infinite matrix. The results are summarized in the introduction. We have found examples to demonstrate the effective existence of each of the types there described.

Part II

THE CONGRUENCES BASED ON ARBITRARY POINTS, PLANES, SPHERES, AND STRAIGHT LINES

7. The general problem

In this part of the paper, the characteristic feature of the analysis is that the initial values α in the formulation of § 5 are functions of a parameter instead of constants.

In the family $\mathfrak{F}$ defined by the differential equations

$$y'' = F(x, y, z, y', z'),$$
$$z'' = G(x, y, z, y', z'),$$

the curve which passes through the point $O(x, y, z)$ in the direction $1:p:q$ is defined by the equations

$$Y = y + p(X-x) + \frac{1}{2}F(X-x)^2 + \cdots,$$
$$Z = z + q(X-x) + \frac{1}{2}G(X-x)^2 + \cdots. \tag{29}$$

If we regard x, y, z as fixed, while p, q vary, these equations represent the congruence of curves of $\mathfrak{F}$ which pass through O, as dependent on the parameters p, q:

$$Y = Y(X; p, q),$$

$$Z = Z(X; p, q).$$

The ω's in this case are the two-rowed determinants in the matrix

$$\begin{vmatrix} Y_p & Z_p & Y'_p & Z'_p \\ Y_q & Z_q & Y'_q & Z'_q \end{vmatrix},$$

and are found by (41) to have for $X = x$, that is, at the fixed point O, the values

(30) $\alpha_1 = 0, \quad \alpha_2 = 0, \quad \alpha_3 = 0, \quad \alpha_4 = 0, \quad \alpha_5 = 0, \quad \alpha_6 = 1.$

We think of p, q as having arbitrary but fixed values, determining thus a definite curve C, which is *any* curve of the congruence based on O. Since the curves passing through an arbitrary point of space are to form a normal congruence, each point of C must serve as the base point of such a congruence. Hence the x for which the ω's take the values (30) is a parameter, capable of continuous variation, *and the α's in* (30) *are therefore to be considered as functions of x having the values* (30) *for all values of x.*

The general problem formulated at the beginning of § 5 is therefore to be modified as follows.

We have

(i) a system of linear differential equations of the first order (independent variable denoted by x),

$$\omega'_i = \sum_j p_{ij}\, \omega_j \qquad (i, j = 1, 2, \ldots, n);$$

(ii) a system of n functions of a parameter x_0,

$$\omega_1^0 = \alpha_1(x_0), \quad \omega_2^0 = \alpha_2(x_0), \ldots, \quad \omega_n^0 = \alpha_n(x_0);$$

(iii) a linear relation

$$A_1\, \omega_1 + A_2\, \omega_2 + \cdots + A_n\, \omega_n = 0$$

whose coefficients are given functions of x.

Our problem is to find the conditions under which the functions ω that satisfy the equations (i), and take for $x = x_0$ the values (ii), obey, for all values of x_0, the linear relation (iii).

We proceed to obtain necessary conditions.

Suppose that the integral curve (in $(n+1)$-space) of the equations (i) which is determined by the point x_0 on the curve of initial points (ii) obeys the relation

$$A_1(x)\,\omega_1(x) + \cdots + A_n(x)\,\omega_n(x) = 0.$$

By differentiation and use of (i) it follows that it must also obey the relation

$$B_1(x)\,\omega_1(x) + \cdots + B_n(x)\,\omega_n(x) = 0,$$

where the B's are derived from the A's by the formula (16_1) (with the number 6 replaced by n).

Repeating this process indefinitely, we are led to the infinite sequence of equations

$$(31)\quad \sum A_i(x)\,\omega_i(x) = 0,\quad \sum B_i(x)\,\omega_i(x) = 0,\quad \sum C_i(x)\,\omega_i(x) = 0, \ldots,$$

where the successive coefficients are derived by the recurrence formulas (16).

The relations (31) are to be identities in x, and must therefore hold in particular for $x = x_0$:

$$(32)\quad \sum A_i(x_0)\,\alpha_i(x_0) = 0,\qquad \sum B_i(x_0)\,\alpha_i(x_0) = 0,$$
$$\sum C_i(x_0)\,\alpha_i(x_0) = 0, \ldots.$$

In our writing $x = x_0$, the letters in (31) change from functions of x to constants, but then become in (32) functions of x_0 through the circumstance that x_0 is capable of taking any value. The equations (32) must be *identities* in x_0, which makes it permissible, as will be done, to differentiate them any number of times.

It is to be seen that in (32) we have *sufficient* as well as necessary conditions for the subsistence of (iii) under the conditions (i) and (ii). For (32) expresses that the derivatives of all orders of the first member of (iii) vanish for $x = x_0$; this must cause the first member of (iii) to be identically zero in x.

We have therefore to undertake the study of the equations (32)—an infinite system of linear equations in the α's. As in Part I, all depends on the rank of the matrix

$$\begin{vmatrix} A_1 \cdots A_n \\ B_1 \cdots B_n \\ C_1 \cdots C_n \\ \cdots\cdots\cdots \end{vmatrix}. \tag{33}$$

It will prevent circumlocution to say that every determinant of higher than the nth order in this matrix is equal to zero.

In the series of sub-matrices of (33),

$$|A|, \quad |AB|, \quad |ABC|, \ldots,$$

let the one of k rows

$$|AB \cdots K|$$

be the first which is equal to zero (in which every determinant resulting from the suppression of columns is equal to zero):

$$|AB \cdots K| = 0, \quad |AB \cdots J| \neq 0. \tag{34}$$

According to the agreement just made, this must happen for $k = n+1$, if not before.

Under (34), the letters K will be linear combinations of $A, B, \ldots, J$:

$$K_i = \varrho_1 A_i + \varrho_2 B_i + \cdots + \varrho_{k-1} J_i,$$

where the ϱ's are functions of x_0. Then by the recurrence formulas (16), the next letters L are linearly composable of those preceding them:

$$L_i = \varrho_1' A_i + (\varrho_2' + \varrho_1) B_1 + \cdots + (\varrho_{k-1}' + \varrho_{k-2}) J_i + \varrho_{k-1} K_i;$$

and the same is seen to be true of every row in (33) from the K row on. This is to say that the first $k-1$ equations of (32),

$$\sum A_i \alpha_i = 0, \quad \sum B^i \alpha_i = 0, \ldots, \quad \sum J_i \alpha_i = 0,$$

will imply all the subsequent ones.

In this way our necessary and sufficient condition—the subsistence of the infinite system (32)—resolves itself into a disjunction of conditions as follows:

In order that the ω's defined by (i) *and* (ii) *satisfy* (iii), *it is necessary and sufficient that one or another of the following $n+1$ mutually exclusive conditions be verified*:

(1) $$|A| = 0;$$

(2) $$|AB| = 0, \qquad |A| \neq 0,$$

$$\sum A_i \alpha_i = 0;$$

(35) (3) $$|ABC| = 0, \qquad |AB| \neq 0,$$

$$\sum A_i \alpha_i = 0, \qquad \sum B_i \alpha_i = 0;$$

. . . . . .

(k) $$|AB \cdots K| = 0, \qquad |AB \cdots J| \neq 0,$$

$$\sum A_i \alpha_i = 0, \ \ldots, \ \sum J_i \alpha_i = 0;$$

and so on to

($n+1$) $$|AB \cdots N| \neq 0,$$

$$\sum A_i \alpha_i = 0, \ \ldots, \ \sum N_i \alpha_i = 0.$$

This is one form of complete solution of the problem formulated at the beginning of this section. But the problem presents a dual aspect.

Subtract from each of the equations (32) the derivative of the preceding. There results the new system

(32′) $$\sum A_i \beta_i = 0, \qquad \sum B_i \beta_i = 0, \qquad \sum C_i \beta_i = 0, \ \ldots,$$

where

(36) $$\beta_i = -\alpha_i' + \sum_{j=1}^{j=n} p_{ij} \alpha_j.$$

Applying the same procedure to the equations (32′), and so on indefinitely, we get the doubly infinite array of equations

$$
\begin{array}{llll}
 & \sum A_i\,\alpha_i = 0, & \sum B_i\,\alpha_i = 0, & \sum C_i\,\alpha_i = 0, \quad \ldots \\
(37) & \sum A_i\,\beta_i = 0, & \sum B_i\,\beta_i = 0, & \sum C_i\,\beta_i = 0, \quad \ldots \\
 & \sum A_i\,\gamma_i = 0, & \sum B_i\,\gamma_i = 0, & \sum C_i\,\gamma_i = 0, \quad \ldots \\
 & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, &
\end{array}
$$

where the γ's are derived from the β's, the δ's from the γ's, etc., by the same formulas which generate the β's from the α's.

Now this array of equations may as well be generated by its first column

$$(38) \qquad \sum A_i\,\alpha_i = 0, \qquad \sum A_i\,\beta_i = 0, \qquad \sum A_i\,\gamma_i = 0, \quad \ldots,$$

as by its first row. For by adding each equation in the first column to the derivative of its predecessor, the second column of equations results, and iteration of this process gives the whole array.

Since the equations (38) thus imply and are implied by the equations (32), they represent, like (32), a necessary and sufficient condition for our problem. And like (32) they are equivalent to a disjunction of mutually exclusive conditions, namely

$$
\begin{array}{ll}
(1) & |\alpha| = 0\,; \\
(2) & |\alpha\beta| = 0, \qquad |\alpha| \neq 0, \\
(39) & \sum A_i\,\alpha_i = 0\,; \\
(3) & |\alpha\beta\gamma| = 0, \qquad |\alpha\beta| \neq 0, \\
 & \sum A_i\,\alpha_i = 0, \qquad \sum A_i\,\beta_i = 0\,;
\end{array}
$$

and so on to

$$
\begin{array}{ll}
(n+1) & |\alpha\beta\cdots\nu| \neq 0, \\
 & \sum A_i\,\alpha_i = 0, \cdots, \sum A_i\,\nu_i = 0.
\end{array}
$$

It is easy to prove through (37) that the case (k) of (35) implies the case $(n-k+2)$ of (39) except for its inequality, and vice versa. For this reason we need never use matrices of order greater than $\frac{1}{2}(n+2)$. Moreover, the elements $\alpha, \beta, \gamma, \ldots$ may be much simpler than $A, B, C, \ldots$. These remarks will be of great advantage to us in the applications that follow.

8. When the base manifold is a point

In the problem of normal congruences, the recurrence formulas for $\beta, \gamma, \delta, \ldots$ are

$$
\begin{aligned}
\beta_1 &= -\alpha_1' && && +\alpha_3 && -\alpha_4, \\
\beta_2 &= -\alpha_2' && +F_z\alpha_1 && +F_p\alpha_2 && +F_q\alpha_3, \\
\beta_3 &= -\alpha_3' && +G_z\alpha_1 && +G_p\alpha_2 && +G_q\alpha_3 && +\alpha_6, \\
\beta_4 &= -\alpha_4' && -F_y\alpha_1 && +F_p\alpha_4 && -F_q\alpha_5 && -\alpha_6, \\
\beta_5 &= -\alpha_5' && +G_y\alpha_1 && -G_p\alpha_4 && +G_q\alpha_5, \\
\beta_6 &= -\alpha_6' && -G_y\alpha_2 && +F_y\alpha_3 && -G_z\alpha_4 && -F_z\alpha_5 && +(F_p+G_q)\alpha_6,
\end{aligned} \tag{40}
$$

the accent denoting the operator $(\partial/\partial x)+p(\partial/\partial y)+q(\partial/\partial z)+F(\partial/\partial p)+G(\partial/\partial q)$. Also, the equation $\sum A_i\alpha_i = 0$ disappears identically, this corresponding to the fact that in every case the congruence starts orthogonally to its base manifold.

Our four problems of the point, plane, sphere, and straight line are each distinguished by a particular form of the functions α.

For the point as base manifold, we have (see (30))

$$
(41)\quad \alpha_1 = 0, \quad \alpha_2 = 0, \quad \alpha_3 = 0, \quad \alpha_4 = 0, \quad \alpha_5 = 0, \quad \alpha_6 = 1.
$$

Calculation gives

$$
(41')\quad \beta_1 = 0, \quad \beta_2 = 0, \quad \beta_3 = 1, \quad \beta_4 = -1, \quad \beta_5 = 0, \quad \beta_6 = F_p + G_q;
$$

$$
\begin{gathered}
\gamma_1 = 2, \quad \gamma_2 = F_q, \quad \gamma_3 = 2G_q + F_p, \quad \gamma_4 = -2F_p - G_q, \\
\gamma_5 = G_p, \quad \gamma_6 = -(F_p+G_q)' + F_y + G_z + (F_p+G_q)^2;
\end{gathered} \tag{41''}
$$

$$
\begin{aligned}
\delta_1 &= 3(F_p+G_q), \\
\delta_2 &= -F_q + 2F_z + 2F_q(F_p+G_q), \\
\delta_3 &= -2F_p' - 3G_q' + F_y + 3G_z + 3G_q(F_p+G_q) + F_p^2 + F_qG_p, \\
\delta_4 &= 2G_q' + 3F_p' - G_z - 3F_y - 3F_p(F_p+G_q) - G_q^2 - F_pG_q, \\
\delta_5 &= -G_p' + 2G_y + 2G_p(F_p+G_q), \\
\delta_6 &= \ldots\ldots\ldots\ldots
\end{aligned} \tag{41'''}
$$

It is easily verified that here, besides $\sum A_i \alpha_i = 0$, we have $\sum B_i \alpha_i = 0$. Hence, by (35), if

$$|AB| = 0,$$

or if

$$|ABC| = 0, \qquad |AB| \neq 0,$$

the congruence determined by an arbitrary point of space will be normal. Thus, in every natural or quasi-natural family the congruence based on an arbitrary point is normal.

We have next to consider

$$(35)\ (4) \qquad \begin{gathered} |ABCD| = 0, \qquad |ABC| \neq 0, \\ \sum A_i \alpha_i = 0, \qquad \sum B_i \alpha_i = 0, \qquad \sum C_i \alpha_i = 0. \end{gathered}$$

These equations written in expanded form would be very complicated on account of the complication of the explicit expressions of the C's and D's; but by the remarks at the end of the preceding section they imply

$$(39)\ (4')^* \qquad \begin{gathered} |\alpha\beta\gamma\delta| = 0, \\ \sum A_i \alpha_i = 0, \qquad \sum A_i \beta_i = 0, \qquad \sum A_i \gamma_i = 0; \end{gathered}$$

and, conversely, (39)(4′) implies (35)(4) except for the condition $|ABC| \neq 0$, since in fact $|\alpha\beta\gamma| \neq 0$ (thus completing the system (39)(4)), indeed

$$(42) \qquad (\alpha_1 \beta_3 \gamma_6) = -2.$$

(39)(4′) gives the following system of four partial differential equations for F and G:

$$(43) \qquad \begin{gathered} F_q(F_p + G_q) + 4F_x - 2(F_{qx} + pF_{qy} + qF_{qz} + FF_{pq} + GF_{qq}) = 0, \\ G_p(F_p + G_q) + 4G_y - 2(G_{px} + pG_{py} + qG_{pz} + FG_{pp} + GG_{pq}) = 0, \\ (1+q^2)F_q + pq(F_p - G_q) - (1+p^2)G_p + 2(pG - qF) = 0, \\ pG_x - qF_x - G_y + F_z = 0, \end{gathered}$$

* By this notation we denote (39)(4) without its inequality, $|\alpha\beta\gamma| \neq 0$.

from the solutions of which we are to exclude those of $|ABC| = 0$. The existence of a solution of (43) which is not simultaneously one of $|ABC| = 0$, that is, not natural or quasi-natural, is proved by the example

$$y'' = y'^2, \qquad z'' = z'^2.$$

There are no further possibilities here than those already enumerated. For the next case is

$$|ABCDE| = 0, \qquad |ABCD| \neq 0,$$
$$\sum A_i \alpha_i = 0, \ \ldots, \ \sum D_i \alpha_i = 0. \tag{35) (5}$$

But these conditions imply (39)(3), which includes

$$|\alpha\beta\gamma| = 0;$$

whereas

$$(\alpha_1 \beta_3 \gamma_6) = -2 \neq 0. \tag{42}$$

9. When the base manifold is a plane

We write the equation of the plane in the form

$$x = -\varphi(y, z) = -(ay + bz + c). \tag{44}$$

The α's are given by (12), where we write $r = s = t = 0$, and find

$$\alpha_1 = 1 + p^2 + q^2, \quad \alpha_2 = qF, \quad \alpha_3 = qG,$$
$$\alpha_4 = -pF, \quad \alpha_5 = pG, \quad \alpha_6 = 0. \tag{45}$$

By applying the formulas (40) the β's are found to have the values

$$\begin{aligned}
\beta_1 &= -pF - qG, \\
\beta_2 &= -qF_x - pqF_y + (1+p^2)F_z - FG, \\
\beta_3 &= -qG_x - pqG_y + (1+p^2)G_z - G^2, \\
\beta_4 &= pF_x - (1+q^2)F_y + pqF_z + F^2, \\
\beta_5 &= -pG_x + (1+q^2)G_y - pqG_z - FG, \\
\beta_6 &= G(qF_y - pF_z) + F(pG_z - qG_y).
\end{aligned} \tag{45'}$$

We omit the calculation of the higher letters.

Passing over the obvious case of natural families, we must next consider

$$(35)\quad (3)\qquad \begin{aligned} |ABC| &= 0, \qquad |AB| \neq 0, \\ \sum A_i \alpha_i &= 0, \qquad \sum B_i \alpha_i = 0. \end{aligned}$$

$\sum A_i \alpha_i = 0$ is satisfied identically, in virtue of the orthogonality of the congruence to its base plane, while $\sum B_i \alpha_i = 0$, which by (22) and (45) is in the present case

$$pG_x - qF_x - G_y + F_z = 0,$$

can be shown to be a *consequence* of the conditions

$$|ABC| = 0, \qquad |AB| \neq 0.$$

For under these conditions, the C's are linearly composable from the A's and B's. Since, by (20), (22), $A_6 = 0$, $B_6 = 0$, we must have $C_6 = 0$. By the sixth formula (21) (write B, C for A, B) it follows that $B_3 = B_4$. But, according to (20), $A_3 = A_4$. Therefore, $C_3 = C_4$. This gives, by further reference to (21) (with B, C for A, B),

$$2B_1 + F_q B_2 + G_q B_3 - F_p B_4 + G_p B_5 + (F_y + G_z) B_6 = 0,$$

which, reduced by (22), is

$$pG_x - qF_x - G_y + F_z = 0.$$

Therefore, *in a quasi-natural family the curves orthogonal to an arbitrary plane form a normal congruence.*

There are three further possibilities here, the conditions for which are given respectively by (4), (5), (6) of (35). We have verified the effective existence of the first of these by the example

$$y'' = \frac{y'(1 + y'^2 + z'^2)}{x + zz'}, \qquad z'' = 0,$$

described, in another connection, in the introduction, but have not been able to settle the existence question for the other two types.*

If we approach the problem from the dual side of the $\alpha, \beta, \gamma, \ldots$, the first case arrived at is

$$|\alpha\beta| = 0.$$

This gives a system of partial differential equations in F and G, whose solution is

$$F = \frac{p(1+p^2+q^2)\,\varphi(\xi)}{\eta\,\psi(\xi)+\zeta\,\varphi(\xi)+\varphi(\xi)\,\psi(\xi)},$$

(46)

$$G = \frac{q(1+p^2+q^2)\,\psi(\xi)}{\eta\,\psi(\xi)+\zeta\,\varphi(\xi)+\varphi(\xi)\,\psi(\xi)},$$

where

$$\xi = x+py+qz, \quad \eta = x+py, \quad \zeta = x+qz.$$

It can be proved that this case is obtainable in the most general way as follows: distribute the ∞^3 planes of space into ∞^2 sets of ∞^1 each, and construct for each set of planes the ∞^2 orthogonal curves. The type (46) is, in other words, characterized by the property that every plane determines a normal congruence in which the other orthogonal surfaces are also planes.

10. When the base manifold is a sphere

For a sphere of radius R with its equation in the form $x = -\varphi(y, z)$,

$$\frac{r}{1+p^2} = \frac{s}{pq} = \frac{t}{1+q^2} = \frac{(1+p^2+q^2)^{1/2}}{R}.$$

Hence, by (12),

$$\alpha_1 = 1+p^2+q^2,$$

(47a)

$$\alpha_2 = qF,$$

$$\alpha_3 = qG + \frac{(1+p^2+q^2)^{3/2}}{R},$$

* A promising approach is to attempt to determine the arbitrary functions φ and ψ in the equations (46) below so as to satisfy the conditions of (35) (5), (35) (6), respectively; but the calculations necessary to carry this through seem too long.

$$\alpha_4 = -pF - \frac{(1+p^2+q^2)^{3/2}}{R},$$

(47b) $$\alpha_5 = pG,$$

$$\alpha_6 = \frac{(1+p^2+q^2)^{1/2}}{R}(pF+qG) + \frac{(1+p^2+q^2)^2}{R^2}.$$

Using the superscripts σ, π, p to denote quantities relative to the sphere, plane, and point, respectively, we see by reference to (41), (41′) and (45) that

$$\alpha^{(\sigma)} = \alpha^{(\pi)} + \lambda\alpha^{(p)} + \mu\beta^{(p)},$$

where

$$\lambda = \frac{(1+p^2+q^2)^{1/2}}{R}(pF+qG) - \frac{(1+p^2+q^2)^{3/2}}{R}(pF+qG) + \frac{(1+p^2+q^2)^2}{R^2},$$

$$\mu = \frac{(1+p^2+q^2)^{3/2}}{R},$$

that is, the *α's for the sphere are linear combinations of the α's for the plane and the α's and β's for the point.* On account of the linear homogeneous character of our entire problem, it follows that if arbitrary points and planes give rise to normal congruences, the same will be true of an arbitrary sphere.

Therefore, by the results of the preceding two sections, *if the family is natural or quasi-natural, every sphere gives rise to a normal congruence.*

We now show that *the natural and quasi-natural families are the only ones in which the curves orthogonal to an arbitrary sphere form a normal congruence.*

For if the normality property subsists with respect to an arbitrary sphere, it must subsist with respect to arbitrary points and planes; we then have by (37)

$$\sum A_i\alpha_i^{(p)} = 0, \quad \sum A_i\beta_i^{(p)} = 0, \quad \sum A_i\gamma_i^{(p)} = 0, \quad \sum A_i\alpha_i^{(\pi)} = 0,$$

$$\sum B_i\alpha_i^{(p)} = 0, \quad \sum B_i\beta_i^{(p)} = 0, \quad \sum B_i\gamma_i^{(p)} = 0, \quad \sum B_i\alpha_i^{(\pi)} = 0,$$

$$\sum C_i\alpha_i^{(p)} = 0, \quad \sum C_i\beta_i^{(p)} = 0, \quad \sum C_i\gamma_i^{(p)} = 0, \quad \sum C_i\alpha_i^{(\pi)} = 0$$

$$(i = 1, \ldots, 6).$$

It follows from this system of linear equations that either

$$|ABC| = 0$$

or

$$|\alpha^{(p)}\ \beta^{(p)}\ \gamma^{(p)}\ \alpha^{(\pi)}| = 0.$$

Under the first condition the family is natural or quasi-natural.

The second gives, by (41), (41′), (41″), (45), the system of partial differential equations

$$(1+p^2+q^2)\,F_q-2qF=0,$$

$$(1+p^2+q^2)\,G_p-2pG=0, \tag{48}$$

$$(1+p^2+q^2)\,F_p-2pF=(1+p^2+q^2)\,G_q-2qG,$$

which together with

$$\sum B_i\,a_i^{(\pi)}=0,$$

whose expansion by (22) and (45) is

$$pG_x-qF_x-G_y+F_z=0, \tag{48′}$$

determine for F and G the forms

$$F=(L_y-pL_x)\,(1+p^2+q^2),$$

$$G=(L_z-qL_x)\,(1+p^2+q^2),$$

that are characteristic of natural families.

This completes the proof of the theorem stated above.

11. When the base manifold is a straight line

Consider the congruence in a system of ∞^4 curves which is determined by orthogonality to an arbitrary straight line

$$y=ax+c,$$

$$z=bx+d.$$

If $1:p:q$ is a direction perpendicular to the line,

$$1+ap+bq=0.$$

The last three equations determine y, z, q in terms of x and p, so that x and p may be used as the parameters of the congruence. The ω's are then the determinants of the matrix

$$\begin{vmatrix} Y_x & Z_x & Y'_x & Z'_x \\ Y_p & Z_p & Y'_p & Z'_p \end{vmatrix}.$$

The initial values of the ω's are found to be

$$\alpha_i = a^2 \alpha_i^{(1)} + ab\,\alpha_i^{(2)} + b^2 \alpha_i^{(3)}, \tag{49}$$

where

$$\alpha_1^{(1)} = 0, \quad \alpha_2^{(1)} = 0, \quad \alpha_3^{(1)} = -(1+p^2), \quad \alpha_4^{(1)} = 0, \quad \alpha_5^{(1)} = pq, \quad \alpha_6^{(1)} = -pF; \tag{49$_1$}$$

$$\alpha_1^{(2)} = 0, \quad \alpha_2^{(2)} = 1+p^2, \quad \alpha_3^{(2)} = -pq, \quad \alpha_4^{(2)} = pq, \quad \alpha_5^{(2)} = 1+q^2, \quad \alpha_6^{(2)} = -qF-pG; \tag{49$_2$}$$

$$\alpha_1^{(3)} = 0, \quad \alpha_2^{(3)} = pq, \quad \alpha_3^{(3)} = 0, \quad \alpha_4^{(3)} = 1+q^2, \quad \alpha_5^{(3)} = 0, \quad \alpha_6^{(3)} = -qG. \tag{49$_3$}$$

A necessary and sufficient condition that the congruence determined by the straight line be normal is given by the infinite system (32), which here is

$$\begin{aligned} &a^2 \textstyle\sum A_i \alpha_i^{(1)} + ab \sum A_i \alpha_i^{(2)} + b^2 \sum A_i \alpha_i^{(3)} = 0, \\ &a^2 \textstyle\sum B_i \alpha_i^{(1)} + ab \sum B_i \alpha_i^{(2)} + b^2 \sum B_i \alpha_i^{(3)} = 0, \\ &a^2 \textstyle\sum C_i \alpha_i^{(1)} + ab \sum C_i \alpha_i^{(2)} + b^2 \sum C_i \alpha_i^{(3)} = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{50}$$

If every straight line gives a normal congruence, these equations must subsist for arbitrary values of the ratio $a:b$; we must then have

$$
(51)\qquad
\begin{array}{lll}
\sum A_i\,\alpha_i^{(1)} = 0, & \sum A_i\,\alpha_i^{(2)} = 0, & \sum A_i\,\alpha_i^{(3)} = 0,\\
\sum B_i\,\alpha_i^{(1)} = 0, & \sum B_i\,\alpha_i^{(2)} = 0, & \sum B_i\,\alpha_i^{(3)} = 0,\\
\sum C_i\,\alpha_i^{(1)} = 0, & \sum C_i\,\alpha_i^{(2)} = 0, & \sum C_i\,\alpha_i^{(3)} = 0,\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & &
\end{array}
$$

Subtracting the derivative of

$$\sum B_i\,\alpha_i^{(1)} = 0$$

from

$$\sum C_i\,\alpha_i^{(1)} = 0,$$

we obtain

$$(52')\qquad \sum B_i\,\beta_i^{(1)} = 0,$$

where the letters $\beta^{(1)}$ are derived from $\alpha^{(1)}$ by the formulas (40), and have the values

$$
(53_1)\qquad
\begin{gathered}
\beta_1^{(1)} = -(1+p^2), \quad \beta_2^{(1)} = -(1+p^2)F_q, \quad \beta_3^{(1)} = pF-(1+p^2)G_q,\\
\beta_4^{(1)} = pF-pqF_q, \quad \beta_5^{(1)} = -qF-pG+pqG_q,\\
\beta_6^{(1)} = F^2+pF_x-F_y+pGF_q-pFG_q.
\end{gathered}
$$

The equations

$$(52)\qquad \sum B_i\,\alpha_i^{(1)} = 0, \quad \sum B_i\,\alpha_i^{(2)} = 0, \quad \sum B_i\,\alpha_i^{(3)} = 0,$$

together with (52′), furnish—when the values of the letters are introduced by means of (22), (49_1), (49_2), (49_3), (53_1)—the same system of partial differential equations which have already appeared as (48), (48′) and which determined for F and G the forms characteristic of natural families.

If every straight line gives rise to a normal congruence, the family must be natural.

Since the equations (50) are quadratic in $a:b$, if they are satisfied for three values of $a:b$, they are satisfied for all. Therefore, *if the straight lines of any triple complex (three lines in each pencil) produce normal congruences the family must be natural.*

Any triple complex of curves will serve as well as straight lines. We omit the proof.

It is sufficient to characterize a family ∞^4 curves as of the natural type that the two complexes of straight lines parallel to the xz and xy planes, respectively, give rise to normal congruences. For then the equations (50) are satisfied for $a = 0$ and for $b = 0$. This gives the two infinite systems of equations

$$\begin{aligned} \sum A_i \alpha_i^{(1)} &= 0, & \sum A_i \alpha_i^{(3)} &= 0, \\ \sum B_i \alpha_i^{(1)} &= 0, & \sum B_i \alpha_i^{(3)} &= 0, \\ \sum C_i \alpha_i^{(1)} &= 0, & \sum C_i \alpha_i^{(3)} &= 0, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & & \cdot\ \cdot\ \cdot\ \cdot\ \cdot & ., \end{aligned}$$

the analysis of which necessitates for F and G the forms typical of natural families.

COLUMBIA UNIVERSITY,
NEW YORK, N.Y.

THE EQUIVALENCE OF CERTAIN REGULAR TRANSFORMATIONS*

BY

L. L. SILVERMAN

1. **Introduction.** The transformation

$$(A) \qquad v(x) = \alpha\, u(x) + \int_0^x k(x,y)\, u(y)\, dy\,,$$

or symbolically, $v = A(u)$, has been studied in special cases by Landau,† and in the general case by the author.‡ It is assumed that $u(x)$ is bounded and integrable, $x_1 \geqq x \geqq 0$, that $k(x,y)$ is integrable in y for each positive x, $0 < \delta \leqq y \leqq x$, and that $\int_0^x |k(x,y)|\, dy$ converges for each positive x. The function $k(x,y)$ will be called the *kernel* and the symbol $[\alpha, k(x,y)]$ the *matrix* of the transformation. The transformation is *regular* if the existence of $\lim_{x\to\infty} u(x)$ implies the existence of $\lim_{x\to\infty} v(x)$ and the equality of the limits. Examples of regular transformations are

$$E\colon\ [1,0], \qquad M\colon\ \left[0, \frac{1}{x}\right],$$

the transformations of Cesàro of order r,

$$(1) \qquad C_r\colon\ \left[0, \frac{r}{x}\left(1 - \frac{y}{x}\right)^{r-1}\right],$$

and the transformations of Hölder of order r,

$$(2) \qquad H_r\colon\ \left[0, \frac{1}{\Gamma(r)}\,\frac{1}{x}\left(\log\frac{x}{y}\right)^{r-1}\right].$$

For $r = 1$, we have $C_1 \equiv H_1 \equiv M$.

* Presented to the Society, April 26, 1919, under the title *Regular transformations of divergent series and integrals.*

† Sächsische Berichte, vol. 65 (1913), p. 131.

‡ These Transactions, vol. 17 (1916), p. 284; Annals of Mathematics, ser. 2, vol. 21 (1919), p. 128.

A sufficient condition for regularity is*

$$\text{(i)} \qquad \int_0^a |k(x,y)|\,dy \text{ converges, } a>0,$$

$$\text{(ii)} \qquad \lim_{x\to\infty}\int_0^a |k(x,y)|\,dy = 0,$$

(3)

$$\text{(iii)} \qquad \int_0^x |k(x,y)|\,dy < N,$$

$$\text{(iv)} \qquad \lim_{x\to\infty}\int_0^x k(x,y)\,dy = 1-\alpha.$$

In a previous paper† the author has discussed a special matrix in terms of a function of a complex variable z, and has obtained conditions for the regularity and equivalence of certain transformations in terms of the corresponding functions of z. In order to avoid the use of complex numbers it is proposed to define the matrix $[\alpha, k(x,y)]$ in terms of the real function $f(x)$ satisfying the following conditions:

$$\text{(i)} \qquad f(x) \text{ is continuous, } 0 \leqq x \leqq 1,$$

$$\text{(ii)} \qquad f'(x) \text{ is continuous, } 0 < x \leqq 1,$$

(4)

$$\text{(iii)} \qquad f(0) = 0,$$

$$\text{(iv)} \qquad \int_0^1 |f'(x)|\,dx \text{ converges.}$$

A function satisfying these conditions will be called *absolutely regular*. We then define

$$\text{(5)} \qquad \alpha = 1-f(1), \qquad k(x,y) = \frac{1}{x}f'\left(\frac{y}{x}\right);$$

* These Transactions, vol. 17 (1916), p. 292.

† Annals of Mathematics, loc. cit., p. 134.

so that solving for the function $f(x)$, we have

$$f(1) = 1 - \alpha, \qquad f(x) = \int_0^x k(1, s)\, ds. \tag{6}$$

We now state the theorem

THEOREM 1. *If the function $f(x)$ is absolutely regular, then the corresponding transformation is regular.*

To prove this theorem, define $\Phi(x) = \int_0^x |f'(t)|\, dt$, $0 < x \leqq 1$; $\Phi(0) = 0$. Then $\Phi(x)$ is continuous, $0 \leqq x \leqq 1$. We now obtain from conditions (3)

$$\int_0^a |k(x, y)|\, dy = \int_0^{a/x} |f'(s)|\, ds = \Phi\left(\frac{a}{x}\right),$$

$$\lim_{x \to \infty} \int_0^a |k(x, y)|\, dy = \lim_{x \to \infty} \Phi\left(\frac{a}{x}\right) = \Phi(0) = 0,$$

$$\int_0^x |k(x, y)|\, dy = \int_0^1 |f'(s)|\, ds = \Phi(1),$$

$$\int_0^x k(x, y)\, dy = \int_0^1 f'(s)\, ds = f(1) = 1 - \alpha.$$

Thus the theorem is proved.

2. **The product transformation.** We now consider three transformations A, B, C corresponding to the absolutely regular functions $f(x)$, $g(x)$, $h(x)$, and inquire as to the form of $h(x)$ when $C = BA$. We have

$$v(x) = \alpha u(x) + \int_0^x k(x, y)\, u(y)\, dy$$

$$= \alpha u(x) + \int_0^1 f'(s)\, u(xs)\, ds; \tag{A}$$

$$w(x) = \beta v(x) + \int_0^1 g'(t) v(xt)\, dt$$

$$(B) \qquad = \alpha\beta u(x) + \beta \int_0^1 f'(s)\, u(xs)\, ds + \alpha \int_0^1 g'(s)\, u(xs)\, ds$$

$$+ \lim_{\varepsilon \to 0} \int_\varepsilon^1 \int_0^1 f'(s)\, g'(t)\, u(xst)\, ds\, dt.$$

Now letting $ts = y$ in the double integral, and then using the abbreviations

$$F(y, t) = \frac{1}{t} g'(t) f'\left(\frac{y}{t}\right) u(xy),$$

$$\Phi(\delta, t) = \int_\delta^t F(y, t)\, dy,$$

the double integral becomes

$$(7) \qquad \lim_{\varepsilon \to 0} \int_\varepsilon^1 \lim_{\delta \to 0} \Phi(\delta, t)\, dt.$$

Now holding ϵ fixed, and letting P be an upper bound of $|u(x)|$, and Q the maximum of $|g'(t)|$ in the interval $(\epsilon, 1)$, we have

$$|\Phi(\delta, t) - \Phi(0, t)| \leqq \int_0^\delta |F(y, t)|\, dy < PQ \int_0^\delta \frac{1}{t} \left| f'\left(\frac{y}{t}\right) \right| dy$$

$$= PQ \int_0^{\delta/t} |f'(s)|\, ds \leqq PQ \int_0^{\delta/\epsilon} |f'(s)|\, ds < \epsilon_1,$$

provided $\delta < \delta_1$, where δ_1 is independent of t. Hence $\lim_{\delta \to 0} \Phi(\delta, t)$ exists uniformly in t, $\epsilon \leqq t \leqq 1$, and

$$\int_{\varepsilon}^{1} \lim_{\delta \to 0} \Phi(\delta, t)\, dt = \lim_{\delta \to 0} \int_{\varepsilon}^{1} \Phi(\delta, t)\, dt$$

so that (7) becomes

$$\lim_{\varepsilon \to 0} \lim_{\delta \to 0} \int_{\varepsilon}^{1} \int_{\delta}^{t} F(y, t)\, dy\, dt$$

$$= \lim_{\delta \to 0} \int_{\delta}^{1} \int_{\delta}^{t} F(y, t)\, dy\, dt + \lim_{\varepsilon \to 0} \lim_{\delta \to 0} \int_{\varepsilon}^{\delta} \int_{\delta}^{t} F(y, t)\, dy\, dt.$$

Now the second term of the last expression is zero; for, letting $\int_{0}^{1} f'(s)\, ds = R$, we have

$$\left| \int_{\varepsilon}^{\delta} \int_{\delta}^{t} F(y, t)\, dy\, dt \right| \leq P \int_{\varepsilon}^{\delta} \int_{\delta}^{t} \left| \frac{g'(t)}{t} f'\left(\frac{y}{t}\right) \right| dy\, dt$$

$$= P \int_{\varepsilon}^{\delta} \int_{\delta/t}^{1} |g'(t) f'(s)|\, ds\, dt < P \int_{\varepsilon}^{\delta} \int_{0}^{1} |g'(t) f'(s)|\, ds\, dt$$

$$= PR \int_{\varepsilon}^{\delta} |g'(t)|\, dt,$$

and the last expression approaches zero*, as δ and ε both approach zero. Hence (7) reduces to

$$\lim_{\delta \to 0} \int_{\delta}^{1} \int_{\delta}^{t} F(y, t)\, dy\, dt = \lim_{\delta \to 0} \int_{\delta}^{1} \int_{y}^{1} F(y, t)\, dt\, dy,$$

* It has been assumed that $\delta \geqq \varepsilon$; the case $\delta < \varepsilon$ would produce only a change in sign in the preceding inequalities and lead to the same conclusion.

and returning to B we may write

$$w(x) = \alpha\beta u(x) + \beta\int_0^1 f'(s)\,u(xs)\,ds + \alpha\int_0^1 g'(s)\,u(xs)\,ds + \int_0^1\left[\int_s^1 \frac{g'(t)}{t} f'\left(\frac{s}{t}\right)dt\right]u(xs)\,ds.$$

Identifying this with

$$w(x) = \gamma u(x) + \int_0^1 h'(s)\,u(xs)\,ds,$$

we have

$$\gamma = \alpha\beta, \text{ or } h(1) = f(1) + g(1) - f(1)\,g(1), \tag{8}$$

$$h'(x) = \beta f'(x) + \alpha g'(x) + \int_x^1 \frac{g'(t)}{t} f'\left(\frac{x}{t}\right)dt. \tag{9}$$

Letting $x/t = s$, we find*

$$h'(x) = \alpha g'(x) + \beta f'(x) + \int_x^1 \frac{f'(s)}{s} g'\left(\frac{x}{s}\right)ds. \tag{10}$$

Hence we see that the transformations A and B are *permutable,*† that is, $AB = BA$. It follows from this that all absolutely regular transformations are consistent.‡

Returning to (9), we have

$$h(x) = \int_0^x h'(t)\,dt = \beta f(x) + \alpha g(x) + \int_0^x\int_t^1 \frac{g'(s)}{s} f'\left(\frac{t}{s}\right)ds\,dt.$$

---

* Cf. Nielsen, *Handbuch der Theorie der Gammafunktion*, p. 119.

† Professor W. A. Hurwitz has proved and communicated to me the theorem that a necessary and sufficient condition that the transformations A and M of § 1 be permutable is that the kernel of A be homogeneous of degree -1.

‡ Hurwitz and Silverman, these Transactions, vol. 18 (1917), p. 7.

Now

$$\int_{t=0}^{x}\int_{s=t}^{1} = \int_{t=0}^{1}\int_{s=t}^{1} - \int_{t=x}^{1}\int_{s=t}^{1} = \int_{s=0}^{1}\int_{t=0}^{s} - \int_{s=x}^{1}\int_{t=x}^{s},$$

whence

$$h(x) = [1-g(1)]\,f(x)+g(x)+\int_{x}^{1} f\left(\frac{x}{s}\right)g'(s)\,ds\,. \tag{11}$$

Similarly

$$h(x) = [1-f(1)]\,g(x)+f(x)+\int_{x}^{1} g\left(\frac{x}{s}\right)f'(s)\,ds\,. \tag{12}$$

Letting $x = 1$ in either (11) or (12) we obtain (8).

Suppose now that $f(x)$ and $g(x)$ are absolutely regular; then $h(x)$ also is absolutely regular. For $h(x)$ clearly satisfies the second and third conditions of (4). As to the fourth condition, we have

$$\int_{\varepsilon}^{1}|h'(x)|\,dx \leqq |\beta|\int_{\varepsilon}^{1}|f'(x)|\,dx+|\alpha|\int_{\varepsilon}^{1}|g'(x)|\,dx + \int_{\varepsilon}^{1}\int_{x}^{1}\frac{1}{t}\left|g'(t)f'\left(\frac{x}{t}\right)\right|dt\,dx\,.$$

Now, interchanging the order of integration in the double integral, making the substitution $x = yt$ in that integral, and letting $R = \int_{0}^{1}|f'(s)|\,ds$, we obtain for the value of the double integral,

$$\int_{\varepsilon}^{1}|g'(t)|\int_{\varepsilon/t}^{1}|f'(y)|\,dy < \int_{\varepsilon}^{1}|g'(t)|\int_{0}^{1}|f'(y)|\,dy\,dt = R\int_{\varepsilon}^{1}|g'(t)|\,dt\,;$$

hence $\int_{0}^{1}|h'(x)|\,dx$ converges. Finally, since $\int_{0}^{x}|h'(t)|\,dt$ is continuous, $0\leqq x\leqq 1$, it follows that $h(x)$ is continuous, $0\leqq x\leqq 1$. Thus $h(x)$ is absolutely regular.

We shall now define an *absolutely regular transformation* as one which corresponds to an absolutely regular function, and collect the results of this section in the following statement:

THEOREM 2. *If the functions $f(x)$, $g(x)$, $h(x)$ correspond to the transformations A, B, C respectively, and if $C = AB$, then $h(x)$ is given in terms of $f(x)$ and $g(x)$ by either of the formulas* (11), (12). *If $f(x)$ and $g(x)$ are absolutely regular, then $h(x)$ also is absolutely regular. All absolutely regular transformations are permutable and hence consistent.*

3. **Inclusion and equivalence.** We shall hereafter denote by the symbol f the transformation corresponding to the function $f(x)$. The transformation f is said to include the transformation g whenever f evaluates the limit of any function evaluated by g, and to the same value; in symbols, $f \geqq g$. If each of two transformations includes the other, they are called equivalent and we write* $f = g$.

We shall now show that in order that the transformation h include the transformation g it is sufficient that there exist a solution $f(x)$ of the integral equation (11) or (12) which is absolutely regular. To see this, let $u(x)$ be any function transformed by h into $v(x)$ and by g into $w(x)$; i. e., $h(u) = v$ and $g(u) = w$. Then $f(w) = fg(u)$. But by hypothesis $fg = h$; hence $f(w) = h(u) = v$. It is now clear that, since f is regular, $\lim_{x\to\infty} v(x) = \xi$ whenever $\lim_{x\to\infty} w(x) = \xi$; in other words $h \geqq g$. Abbreviating equations (11), (12), by the symbols

$$h(x) = g(x) \circ f(x), \tag{11}$$

$$h(x) = f(x) \circ g(x), \tag{12}$$

we may say

THEOREM 3. *A sufficient condition that the transformation f include the transformation g is that there exist an absolutely regular solution $p(x)$ of the integral equation $f(x) = p(x) \circ g(x)$.*

COROLLARY 1. *A sufficient condition that the transformations f and g be equivalent is that there exist two absolutely regular functions $p(x), q(x)$ satisfying the equations $f(x) = p(x) \circ g(x)$, $g(x) = q(x) \circ f(x)$.*

COROLLARY 2. *A sufficient condition that the transformation f be equivalent to the identical transformation E is that there exist an absolutely regular solution of the equation $f(x) \circ p(x) = 0$.*

4. **Examples.**

(i) The function corresponding to the transformation of Cesàro of order r is $C_r(x) = 1-(1-x)^r$.

This follows directly from (1) and (6).

* The notation $f \equiv g$ will indicate the identity of f and g.

(ii) The function corresponding to the transformation of Hölder of order r is

$$h_r(x) = \frac{1}{\Gamma(r)} \int_0^x \left(\log \frac{1}{t}\right)^{r-1} dt.$$

This follows directly from (2) and (6).

(iii) The transformations of Cesàro and Hölder of order 2 are* equivalent.

To prove this we shall prove the existence of two absolutely regular functions $p(x)$, $q(x)$ satisfying the conditions $c_2(x) = h_2(x) \circ p(x)$, $h_2(x) = C_2(x) \circ q(x)$. To determine $p(x)$ we have the equation

$$1-(1-x)^2 = x - x\log x + \int_x^1 p\left(\frac{x}{s}\right) \log \frac{1}{s}\, ds.$$

Making in the integral the substitution $t = x/s$, dividing by x, differentiating both sides with respect to x, multiplying by x, and again differentiating both sides with respect to x, we obtain $p(x) = -x^2$, which is obviously absolutely regular.

To determine $q(x)$ we have the equation

$$x - x\log x = 1-(1-x)^2 + 2\int_x^1 q\left(\frac{x}{s}\right)(1-s)\, ds.$$

Making the substitution $t = x/s$, dividing by x and differentiating both sides twice with respect to x we obtain $q(x) = x/2$, which is obviously absolutely regular.

(iv) The functions

$$f(x) = ax^n, \quad g(x) = \frac{a}{a-1} x^{\frac{n}{1-a}}, \quad n>0, \quad a<1,$$

correspond to transformations equivalent to convergence and hence equivalent to each other. For

$$f(x) \circ g(x) = \frac{a}{1-a} x^n - \frac{a}{1-a} x^{n/(1-a)} - \frac{na^2}{(1-a)^2} \int_x^1 \left(\frac{x}{s}\right)^n s^{n/(1-a)-1}\, ds$$

$$= 0.$$

* The method here employed can be used also to deduce the well known theorem concerning the equivalence of these transformations for any integral order r. The case of non-integral order will be deduced more readily from a test for equivalence to be published in a later paper.

Now if $n>0$, $a<1$, both functions are absolutely regular, hence our statement is proved.

(v) The transformation* $\alpha E+(1-\alpha)M$ is reversible (equivalent to convergence) when $\alpha>0$.

This follows directly from example (iv), putting $n=1$, $\alpha=1-a$.

5. **Transformation of sequences.** The results which we have obtained in the preceding sections for the transformations of functions hold also for transformations of sequences. But inasmuch as a very general treatment of this subject has recently been given by Hausdorff,† we shall consider it only briefly. It should be stated, however, that while Hausdorff devotes some attention to the question of inclusion and equivalence, he does not consider, as we do, the general question of the inclusion and equivalence of absolutely regular transformations.

Let us consider the transformation

$$y_n=\sum_{k=1}^{n}a_{nk}x_k \qquad (n=1,2,3,\ldots) \tag{13}$$

where the matrix (a_{nk}), $k\leqq n$, is a triangular matrix. A sufficient condition that two such transformations be permutable is‡ that their matrices be expressible in the form

$$a_{nk}=\sum_{h=k}^{n}(-1)^{h-k}\frac{(n-1)!}{(n-h)!\,(h-k)!\,(k-1)!}a_{hh}. \tag{14}$$

We shall restrict ourselves to matrices of this form and we shall define a_{nn} in terms of an absolutely regular function $f(x)$,

$$a_{nn}=1-f(1)+\int_0^1 t^{n-1}f'(t)\,dt. \tag{15}$$

The transformation so defined is regular§ and will be called absolutely regular. Hausdorff considers the more general case where $f(x)$ is a function with

* I. Schur, Mathematische Annalen, vol. 74 (1913), p. 447; Landau, Sächsische Berichte, vol. 65, p. 131.

† Mathematische Zeitschrift, vol. 9 (1921), p. 74 and p. 280.

‡ These Transactions, vol. 18 (1917), p. 7.

§ Hausdorff, loc. cit., p. 84.

limited total fluctuation and $\int_0 t^{n-1}\,df(t)$ is a Stieltjes integral. To obtain the function $h(x)$ corresponding to the product (c_{nk}) of two transformations (a_{nk}) and (b_{nk}) we need only compute c_{nn}. If $f(x)$, $g(x)$ correspond to (a_{nk}), (b_{nk}), we find $c_{nn} = a_{nn}\,b_{nn}$, or

$$\gamma + \int_0^1 t^{n-1} h'(t)\,dt = \left(\alpha + \int_0^1 t^{n-1} f'(t)\,dt\right)\left(\beta + \int_0^1 t^{n-1} g'(t)\,dt\right), \tag{16}$$

where $\alpha = 1-f(1)$, $\beta = 1-g(1)$, $\gamma = 1-h(1)$. Hence we must determine a function $h(x)$ such that

$$\gamma = \alpha\beta, \text{ i.e., } h(1) = f(1)+g(1)-f(1)\,g(1), \tag{17}$$

and

$$\int_0^1 t^{n-1} h'(t)\,dt = \alpha\int_0^1 t^{n-1} g'(t)\,dt + \beta\int_0^1 t^{n-1} f'(t)\,dt + \int_0^1\int_0^1 s^{n-1}\,t^{n-1} f'(s)\,g'(t)\,ds\,dt. \tag{18}$$

But such a function $h(x)$ is precisely $h(x)$ as given by formulas (11), (12). For, multiplying (10) by x^{n-1} and integrating from 0 to 1, interchanging the order of integration* in the resulting double integral, and then in that integral making the substitution $t = x/s$, we obtain precisely (18). Furthermore, letting $x = 1$ in (12), we see that (17) also is satisfied. Finally (18) cannot have more than one solution, for if $h_1(x)$ is a second solution, we should have

$$\int_0^1 t^{n-1}\,(h'(t)-h_1'(t)\,dt = 0;$$

* The justification for the interchange of the order of integration can be made as on page 105.

and hence, by a theorem due to Lerch* and Stieltjes,† $h_1'(t) = h'(t)$. And since $h_1(0) = h(0) = 0$, it follows that $h_1(x) = h(x)$.

It has now been shown that the results of articles 1–3 hold also for sequences.

6. **Conclusion.** We have obtained two results in this paper. First we have defined the matrix of a transformation for both sequences and functions in terms of an absolutely regular function, and have then obtained the same formula for the function corresponding to the product of two transformations in the case of both sequences and functions. Secondly, we have obtained definite criteria (the same criteria for both sequences and functions) for the inclusion and equivalence of absolutely regular transformations. In another paper we shall consider some further criteria for the inclusion and equivalence of these transformations.

---

*Rozpravy ceské Akademie, 2d class, vol. 1 (1892), vol. 2 (1893).

† *Correspondence d'Hermite et de Stieltjes*, vol. 2, pp. 337–339.

DARTMOUTH COLLEGE,
HANOVER, N. H.

MACLAURIN EXPANSION OF THE INTERPOLATION POLYNOMIAL DETERMINED BY $2n+1$ EVENLY SPACED POINTS*

BY

GEORGE RUTLEDGE

1. **Introduction.** In a recent paper† by the writer, the polynomial determined by the $2n+1$ points

$$(1) \qquad (-nh, y_{-n}), \ldots, (-h, y_{-1}), (0, y_0), (h, y_1), \ldots (nh, y_n)$$

is written down in terms of the determinant

$$(2) \qquad D = \begin{vmatrix} 1 & 2^2 & 3^2 & \cdots & n^2 \\ 1 & 2^4 & 3^4 & \cdots & n^4 \\ 1 & 2^6 & 3^6 & \cdots & n^6 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & 2^{2n} & 3^{2n} & \cdots & n^{2n} \end{vmatrix} = n!\,3!\,5! \cdots (2n-1)!$$

* Presented to the Society, February 24, 1923.

† Journal of Mathematics and Physics of the Massachusetts Institute of Technology, vol. 2 (1922), p. 47. In this paper the cofactors A_{ij} are explicitly expressed as follows:

$$\frac{A_{ij}}{2D} = (-1)^{i+j} \frac{\binom{2n}{n-j}}{(2n)!} \sum^{(j)} r_1^2 r_2^2 \cdots r_{n-i}^2,$$

where $\sum^{(j)} r_1^2 r_2^2 \cdots r_{n-i}^2$ represents the sum of the squares of the $\binom{n-1}{n-i}$ products of the first n integers excepting j, taken $n-i$ at a time, $i \neq n$. For $i = n$, $\sum^{(j)} r_1^2 r_2^2 \cdots r_{n-i}^2$ is unity by definition.

and the cofactors A_{ij} of its elements j^{2i}, as follows:

$$
\begin{aligned}
P^{[2n]}(x) &= y_0 \\
&+ \left\{ \frac{A_{11}}{2D}(y_1 - y_{-1}) + \frac{2A_{12}}{2D}(y_2 - y_{-2}) + \cdots + \frac{nA_{1n}}{2D}(y_n - y_{-n}) \right\} \frac{x}{h} \\
&+ \left\{ \frac{A_{11}}{2D}(y_1 + y_{-1} - 2y_0) + \frac{A_{12}}{2D}(y_2 + y_{-2} - 2y_0) + \cdots \right. \\
&\qquad \left. \cdots + \frac{A_{1n}}{2D}(y_n + y_{-n} - 2y_0) \right\} \frac{x^2}{h^2} \\
&\qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \qquad (3) \\
&+ \left\{ \frac{A_{n1}}{2D}(y_1 - y_{-1}) + \frac{2A_{n2}}{2D}(y_2 - y_{-2}) + \cdots + \frac{nA_{nn}}{2D}(y_n - y_{-n}) \right\} \frac{x^{2n-1}}{h^{2n-1}} \\
&+ \left\{ \frac{A_{n1}}{2D}(y_1 + y_{-1} - 2y_0) + \frac{A_{n2}}{2D}(y_2 + y_{-2} - 2y_0) + \cdots \right. \\
&\qquad \left. \cdots + \frac{A_{nn}}{2D}(y_n + y_{-n} - 2y_0) \right\} \frac{x^{2n}}{h^{2n}}.
\end{aligned}
$$

This formula is comparable with the Lagrange formula in ease of verification (independently of evaluation of the determinant D and its cofactors).

The object of the present paper is to express the coefficient of x^m in (3) in terms of mth difference quotients, thus obtaining a form of (3) which suggests Maclaurin's series as a limiting case under proper conditions.

2. **Transformation of the coefficients.** The scheme of differences to be employed is the following:

$$
\begin{array}{lccccccccc}
y_{-3} & & & & & & & & & \\
 & \Delta'_{-3} & & & & & & & & \\
y_{-2} & & \Delta''_{-2} & & & & & & & \\
 & \Delta'_{-2} & & \Delta'''_{-2} & & & & & & \\
y_{-1} & & \Delta''_{-1} & & \Delta^{IV}_{-1} & & & & & \\
 & \Delta'_{-1} & & \Delta'''_{-1} & & \Delta^{V}_{-1} & & & & \\
y_0 & & \Delta''_0 & & \Delta^{IV}_0 & & \Delta^{VI}_0 & & \cdots, & (4) \\
 & \Delta'_1 & & \Delta'''_1 & & \Delta^{V}_1 & & & & \\
y_1 & & \Delta''_1 & & \Delta^{IV}_1 & & & & & \\
 & \Delta'_2 & & \Delta'''_2 & & & & & & \\
y_2 & & \Delta''_2 & & & & & & & \\
 & \Delta'_3 & & & & & & & & \\
y_3 & & & & & & & & &
\end{array}
$$

where $y_3 - y_2 = \Delta'_2$, $\Delta'_3 - \Delta'_2 = \Delta''_2$, etc.

Between the differences of orders one and three, and between the differences of orders two and four, we have, respectively, the following relations:

$$\begin{aligned}
&\Delta'_{-1}+\Delta'_1 = (\Delta'_{-1}+\Delta'_1),\\
&\Delta'_{-2}+\Delta'_2 = (\Delta'_{-1}+\Delta'_1)+ (\Delta'''_{-1}+\Delta'''_1),\\
&\Delta'_{-3}+\Delta'_3 = (\Delta'_{-1}+\Delta'_1)+2(\Delta'''_{-1}+\Delta'''_1)+ (\Delta'''_{-2}+\Delta'''_2),\\
&\Delta'_{-4}+\Delta'_4 = (\Delta'_{-1}+\Delta'_1)+3(\Delta'''_{-1}+\Delta'''_1)+2(\Delta'''_{-2}+\Delta'''_2)+(\Delta'''_{-3}+\Delta'''_3),\\
&\quad\cdots\cdots\cdots

The coefficients of (7) are now transformed by successive application of (5) and (6) (and the generalization of these relations) until the coefficient of x^{2i-1} contains differences of order $2i-1$ and the coefficient of x^{2i} contains differences of order $2i$. *It will then appear that the differences of lower order, though formally involved, have actually been eliminated.*

By use of (5) and the identity

$$(8)\quad \binom{n}{k}\cdot 1+\binom{n-1}{k}\cdot 2+\binom{n-2}{k}\cdot 3+\cdots+\binom{k}{k}\cdot(n-k+1)=\binom{n+2}{k+2},$$

we may reduce the expression

$$(9)\quad \left[\binom{j-1}{0}(\Delta'_{-1}+\Delta'_1)+\binom{j-2}{0}(\Delta'_{-2}+\Delta'_2)+\cdots+\binom{0}{0}(\Delta'_{-j}+\Delta'_j)\right]$$

to the form

$$(10)\quad \left[\binom{j}{2}(\Delta'''_{-1}+\Delta'''_1)+\binom{j-1}{2}(\Delta'''_{-2}+\Delta'''_2)+\cdots+\binom{2}{2}(\Delta'''_{-j+1}+\Delta'''_{j-1})\right],$$

except for a term in $(\Delta'_{-1}+\Delta'_1)$ which by combination with similar terms for $j=1,2,3,\ldots,n$ is eliminated from the coefficients of $x^3, x^5, \ldots$, as stated above, and as will presently appear.

Repetition of this process yields from (10)

$$(11)\quad \left[\binom{j+1}{4}(\Delta^{\mathrm{v}}_{-1}+\Delta^{\mathrm{v}}_1)+\binom{j}{4}(\Delta^{\mathrm{v}}_{-2}+\Delta^{\mathrm{v}}_2)+\cdots+\binom{4}{4}(\Delta^{\mathrm{v}}_{-j+2}+\Delta^{\mathrm{v}}_{j-2})\right],$$

and so on, the final expression in the sequence being

$$(12)\quad \left[\binom{2j-2}{2j-2}(\Delta^{[2j-1]}_{-1}+\Delta^{[2j-1]}_1)\right].$$

In like manner by use of (6) and the identity (8) we may reduce the expression

$$(13)\quad \left[\binom{j}{1}(\Delta''_0)+\binom{j-1}{1}(\Delta''_{-1}+\Delta''_1)+\cdots+\binom{1}{1}(\Delta''_{-j+1}+\Delta''_{j-1})\right]$$

to the forms

$$(14)\quad \left[\binom{j+1}{3}(\Delta_0^{IV})+\binom{j}{3}(\Delta_{-1}^{IV}+\Delta_1^{IV})+\cdots+\binom{3}{3}(\Delta_{-j+2}^{IV}+\Delta_{j-2}^{IV})\right],$$

$$(15)\quad \left[\binom{j+2}{5}(\Delta_0^{VI})+\binom{j+1}{5}(\Delta_{-1}^{VI}+\Delta_1^{VI})+\cdots+\binom{5}{5}(\Delta_{-j+3}^{VI}+\Delta_{j-3}^{VI})\right],$$

and so on, with exceptions of the same nature as in the case of the expression (9). The final expression in the sequence is

$$(16)\quad \left[\binom{2j-1}{2j-1}(\Delta_0^{[2j]})\right].$$

3. **Enumeration of the differences.** By use of the identities

$$(17)\quad \binom{n}{k}+\binom{n-1}{k}+\binom{n-2}{k}+\cdots+\binom{k}{k}=\binom{n+1}{k+1}$$

and

$$(18)\quad \binom{n}{k}+2\binom{n-1}{k}+2\binom{n-2}{k}+\cdots+2\binom{k}{k}=\frac{2n-k+1}{k+1}\binom{n}{k},$$

we find that the expressions (9), (10), (11), ... contain, respectively,

$$(19)\quad 2\binom{j}{1},\qquad 2\binom{j+1}{3},\qquad 2\binom{j+2}{5},\ \cdots$$

differences, and that the expressions (13), (14), 15), ... contain, respectively,

$$(20)\quad \frac{j}{1}\binom{j}{1},\qquad \frac{j}{2}\binom{j+1}{3},\qquad \frac{j}{3}\binom{j+2}{5},\cdots$$

differences.

Consequently, if we replace the expressions in the brackets in (7) by the reduced expressions in the higher differences, (7) becomes (apart from the omitted differences of lower order) *with respect only to the number of differences involved*

$$y_0+\sum_{i=1}^{n}\sum_{j=i}^{n}\frac{A_{ij}}{D}\left[j\binom{j+i-1}{2i-1}\Delta^{[2i-1]}\text{'s}\right]\frac{x^{2i-1}}{h^{2i-1}}$$

(21)

$$+\sum_{i=1}^{n}\sum_{j=i}^{n}\frac{A_{ij}}{D}\left[\frac{j}{2i}\binom{j+i-1}{2i-1}\Delta^{[2i]}\text{'s}\right]\frac{x^{2i}}{h^{2i}},$$

where, because of the termination, for $i=j$, of the sequences of expressions (9), (10), (11), ..., and (13), (14), (15), ..., we have $j \geqq i$.

We shall now show that

$$\sum_{j=i}^{n}\frac{A_{ij}}{D}\,j\binom{j+i-1}{2i-1}=\frac{1}{(2i-1)!} \tag{22}$$

and hence that

$$\sum_{j=i}^{n}\frac{A_{ij}}{D}\,\frac{j}{2i}\binom{j+i-1}{2i-1}=\frac{1}{(2i)!}; \tag{23}$$

also that

$$\sum_{j=k}^{n}\frac{A_{ij}}{D}\,j\binom{j+k-1}{2k-1}=0,\ k<i, \tag{24}$$

and hence that

$$\sum_{j=k}^{n}\frac{A_{ij}}{D}\,\frac{j}{2k}\binom{j+k-1}{2k-1}=0,\ k<i. \tag{25}$$

We shall find it convenient to introduce the notation ($j \geqq i$)

$$a_{ij}=j(2i-1)!\binom{j+i-1}{2i-1}=j^2(j^2-1)(j^2-2^2)\cdots(j^2-(i-1)^2). \tag{26}$$

In order to establish (22), (23), (24), (25), we have then to prove

$$\sum_{j=i}^{n}a_{ij}A_{ij}=D, \tag{27}$$

and

$$\sum_{j=k}^{n}a_{kj}A_{ij}=0,\ k<i. \tag{28}$$

But the determinant D is readily written in the form

$$(29)\quad D = \begin{vmatrix} 1 & 2^2 & 3^2 & 4^2 & \cdots \\ 0 & 2^2(2^2-1) & 3^2(3^2-1) & 4^2(4^2-1) & \cdots \\ 0 & 0 & 3^2(3^2-1)(3^2-2^2) & 4^2(4^2-1)(4^2-2^2) & \cdots \\ 0 & 0 & 0 & 4^2(4^2-1)(4^2-2^2)(4^2-3^2) & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \end{vmatrix},$$

the nth row of this form being obtained by multiplying the rows of (2) in order by the coefficients of $x^2, x^4, x^6, \ldots, x^{2n}$ in the polynomial $x^2(x^2-1)(x^2-2^2)\cdots(x^2-(n-1)^2)$ and adding. The remaining rows are obtained in like manner. If only the first i rows of (2) are so transformed the cofactors of elements in rows i to n inclusive are the same in (29) as in (2). From this the truth of (27) and (28) is at once evident. From (24) and (25), which are now established, it is clear that differences of lower order are eliminated by the transformations of Section 2, as there predicted, and from this fact in combination with (22) and (23) we have the following result:

THEOREM. *The coefficient of x^{2n} in the polynomial determined by the points* (1) *is the product of the reciprocal of $n!$ and a weighted average of difference quotients of order n.*

4. **Explicit expansion of $P^{[2n]}(x)$ in terms of difference quotients.** In the light of the foregoing theorem the explicit expansion of the polynomial $P^{[2n]}(x)$ in the manner specified by the theorem becomes important. This expansion results from the transformations of Section 2, and is as follows:

$$(30)\quad \begin{aligned} P^{[2n]}(x) = y_0 + \sum_{i=1}^{n}\sum_{j=i}^{n}\frac{jA_{ij}}{2D}\Bigg[&\binom{j+i-2}{2i-2}(\Delta_{-1}^{[2i-1]}+\Delta_1^{[2i-1]}) \\ &+\binom{j+i-3}{2i-2}(\Delta_{-2}^{[2i-1]}+\Delta_2^{[2i-1]})+\cdots+\binom{2i-2}{2i-2}(\Delta_{-j+i-1}^{[2i-1]}+\Delta_{j-i+1}^{[2i-1]})\Bigg]\frac{x^{2i-1}}{h^{2i-1}} \\ +\sum_{i=1}^{n}\sum_{j=i}^{n}\frac{A_{ij}}{2D}\Bigg[&\binom{j+i-1}{2i-1}(\Delta_0^{[2i]})+\binom{j+i-2}{2i-1}(\Delta_{-1}^{[2i]}+\Delta_1^{[2i]}) \\ &+\cdots+\binom{2i-1}{2i-1}(\Delta_{-j+i}^{[2i]}+\Delta_{j-i}^{[2i]})\Bigg]\frac{x^{2i}}{h^{2i}}. \end{aligned}$$

Combination of (26) and (30) now enables us to write $P^{[2n]}(x)$ in the form*

$$
\begin{aligned}
P^{[2n]}(x) &= y_0 \\
&+ \sum_{i=1}^{n} \frac{1}{(2i-1)!} \sum_{k=1}^{n-i+1} \frac{\Delta_{-k}^{[2i-1]} + \Delta_k^{[2i-1]}}{2h^{2i-1}} \cdot \frac{\sum_{j=i-1+k}^{n} j\binom{j+i-1-k}{2i-2} A_{ij}}{\sum_{j=i}^{n} j\binom{j+i-1}{2i-1} A_{ij}} x^{2i-1} \\
&+ \sum_{i=1}^{n} \frac{1}{(2i)!} \sum_{k=0}^{n-i} \frac{\Delta_{-k}^{[2i]} + \Delta_k^{[2i]}}{2h^{2i}} \cdot \frac{\sum_{j=i+k}^{n} 2i\binom{j+i-1-k}{2i-1} A_{ij}}{\sum_{j=i}^{n} j\binom{j+i-1}{2i-1} A_{ij}} x^{2i}.
\end{aligned} \tag{31}
$$

Formula (31) presents explicitly the weightings of the various difference quotients in the weighted averages mentioned in the theorem of Section 3.

For $n = 1, 2, 3, 4$ we have the following numerical results:

The polynomial determined by three points ($n = 1$):

$$
P^{[2]}(x) = y_0 + \frac{\Delta'_{-1} + \Delta'_1}{2h} x + \frac{1}{2!} \frac{\Delta''_0}{h^2} x^2. \tag{32}
$$

The polynomial determined by five points ($n = 2$):

$$
\begin{aligned}
P^{[4]}(x) = y_0 &+ \left\{ \frac{7}{6} \frac{\Delta'_{-1} + \Delta'_1}{2h} - \frac{1}{6} \frac{\Delta'_{-2} + \Delta'_2}{2h} \right\} x \\
&+ \frac{1}{2!} \left\{ \frac{7}{6} \frac{\Delta''_0}{h^2} - \frac{1}{6} \frac{\Delta''_{-1} + \Delta''_1}{2h^2} \right\} x^2 \\
&+ \frac{1}{3!} \left\{ \frac{\Delta'''_{-1} + \Delta'''_1}{2h^3} \right\} x^3 + \frac{1}{4!} \left\{ \frac{\Delta_0^{\mathrm{IV}}}{h^4} \right\} x^4.
\end{aligned} \tag{33}
$$

The polynomial determined by seven points ($n = 3$):

$$
\begin{aligned}
P^{[6]}(x) = y_0 &+ \left\{ \frac{111}{90} \frac{\Delta'_{-1} + \Delta'_1}{2h} - \frac{24}{90} \frac{\Delta'_{-2} + \Delta'_2}{2h} + \frac{3}{90} \frac{\Delta'_{-3} + \Delta'_3}{2h} \right\} x \\
&+ \frac{1}{2!} \left\{ \frac{111}{90} \frac{\Delta''_0}{h^2} - \frac{23}{90} \frac{\Delta''_{-1} + \Delta''_1}{2h^2} + \frac{2}{90} \frac{\Delta''_{-2} + \Delta''_2}{2h^2} \right\} x^2
\end{aligned}
$$

* For $k = 0$, $\Delta_{-k}^{[2i]} = 0$, by definition.

$$+\frac{1}{3!}\left\{\frac{5}{4}\,\frac{\Delta'''_{-1}+\Delta'''_{1}}{2h^3}-\frac{1}{4}\,\frac{\Delta'''_{-2}+\Delta'''_{2}}{2h^3}\right\}x^3$$

$$+\frac{1}{4!}\left\{\frac{4}{3}\,\frac{\Delta^{\mathrm{IV}}_{0}}{h^4}-\frac{1}{3}\,\frac{\Delta^{\mathrm{IV}}_{-1}+\Delta^{\mathrm{IV}}_{1}}{2h^4}\right\}x^4 \tag{34}$$

$$+\frac{1}{5!}\left\{\frac{\Delta^{\mathrm{V}}_{-1}+\Delta^{\mathrm{V}}_{1}}{2h^5}\right\}x^5+\frac{1}{6!}\left\{\frac{\Delta^{\mathrm{VI}}_{0}}{h^6}\right\}x^6.$$

The polynomial determined by nine points ($n = 4$):

$$P^{[8]}(x) = y_0+\left\{\frac{3198}{2520}\,\frac{\Delta'_{-1}+\Delta'_{1}}{2h}-\frac{834}{2520}\,\frac{\Delta'_{-2}+\Delta'_{2}}{2h}+\frac{174}{2520}\,\frac{\Delta'_{-3}+\Delta'_{3}}{2h}-\frac{18}{2520}\,\frac{\Delta'_{-4}+\Delta'_{4}}{2h}\right\}x$$

$$+\frac{1}{2!}\left\{\frac{3198}{2520}\,\frac{\Delta''_{0}}{h^2}-\frac{779}{2520}\,\frac{\Delta''_{-1}+\Delta''_{1}}{2h^2}+\frac{110}{2520}\,\frac{\Delta''_{-2}+\Delta''_{2}}{2h^2}-\frac{9}{2520}\,\frac{\Delta''_{-3}+\Delta''_{3}}{2h^2}\right\}x^2$$

$$+\frac{1}{3!}\left\{\frac{164}{120}\,\frac{\Delta'''_{-1}+\Delta'''_{1}}{2h^3}-\frac{51}{120}\,\frac{\Delta'''_{-2}+\Delta'''_{2}}{2h^3}+\frac{7}{120}\,\frac{\Delta'''_{-3}+\Delta'''_{3}}{2h^3}\right\}x^3 \tag{35}$$

$$+\frac{1}{4!}\left\{\frac{181}{120}\,\frac{\Delta^{\mathrm{IV}}_{0}}{h^4}-\frac{68}{120}\,\frac{\Delta^{\mathrm{IV}}_{-1}+\Delta^{\mathrm{IV}}_{1}}{2h^4}+\frac{7}{120}\,\frac{\Delta^{\mathrm{IV}}_{-2}+\Delta^{\mathrm{IV}}_{2}}{2h^4}\right\}x^4$$

$$+\frac{1}{5!}\left\{\frac{4}{3}\,\frac{\Delta^{\mathrm{V}}_{-1}+\Delta^{\mathrm{V}}_{1}}{2h^5}-\frac{1}{3}\,\frac{\Delta^{\mathrm{V}}_{-2}+\Delta^{\mathrm{V}}_{2}}{2h^5}\right\}x^5$$

$$+\frac{1}{6!}\left\{\frac{3}{2}\,\frac{\Delta^{\mathrm{VI}}_{0}}{h^6}-\frac{1}{2}\,\frac{\Delta^{\mathrm{VI}}_{-1}+\Delta^{\mathrm{VI}}_{1}}{2h^6}\right\}x^6$$

$$+\frac{1}{7!}\left\{\frac{\Delta^{\mathrm{VII}}_{-1}+\Delta^{\mathrm{VII}}_{1}}{2h^7}\right\}x^7+\frac{1}{8!}\left\{\frac{\Delta^{\mathrm{VIII}}_{0}}{h^8}\right\}x^8.$$

The weightings of the mean difference quotients of the first and second orders in the weighted averages involved in the coefficients of x and x^2 are of particular interest, and for low degrees are of practical importance. We therefore tabulate these weightings for degrees 2, 4, 6, 8, 10, the weightings in any given column being in order of numerically ascending subscripts of the mean difference quotients:

Weightings of first Difference Quotients

(36)

Degree 2	Degree 4	Degree 6	Degree 8	Degree 10
1.000 000	1.166 666	1.233 333	1.269 047 61	1.291 269 84
	− 0.166 666	− 0.266 666	− 0.330 952 38	− 0.375 396 82
		0.033 333	0.069 047 61	0.100 793 65
			− 0.007 142 85	− 0.018 253 96
				0.001 587 30

Weightings of second Difference Quotients

(37)

Degree 2	Degree 4	Degree 6	Degree 8	Degree 10
1.000 000	1.166 666	1.233 333	1.269 047 61	1.291 269 84
	− 0.166 666	− 0.255 555	− 0.309 126 984	− 0.344 682 539
		0.022 222	0.043 650 79	0.061 428 57
			− 0.003 571 428	− 0.008 650 793
				0.000 634 920

By means of (36) and (37) the first and second derivatives of $P^{[2n]}(x)$, $n = 1, 2, 3, 4, 5$, at the point $(0, y_0)$ are quickly determined without computing differences of higher order than that of the desired derivative. The values given in (36) and (37) are readily checked by taking for $y_{-n}, \ldots, y_{-1}, y_0, y_1, \ldots, y_n$ the values, for $x = -n, \ldots, -1, 0, 1, \ldots, n$, respectively, of any polynomial (for example, x^m) of degree less than or equal to $2n$.

5. **Derivation of Stirling's form of** $P^{[2n]}(x)$. The polynomial $P^{[2n]}(x)$ is derived in the foregoing by methods in no way dependent on the classic formulas of Lagrange and Stirling. Stirling's formula, on the other hand, is very readily derived from (31).

We shall use the customary notation

$$\frac{\Delta'_{-1} + \Delta'_1}{2} = \mu \Delta'_0, \qquad \frac{\Delta'''_{-1} + \Delta'''_1}{2} = \mu \Delta'''_0, \ldots. \tag{38}$$

From (32) we have

$$P^{[2]}(x) = y_0 + \mu \Delta'_0 \frac{x}{h} + \frac{1}{2!} \Delta''_0 \frac{x^2}{h^2}. \tag{39}$$

If to (39) we add a polynomial of degree 4 which has for its terms of the third and fourth degree the terms of the third and fourth degree of $P^{[4]}(x)$, as specified by (33), and which vanishes for $x/h = -1, 0, 1$, the difference between the polynomial thus obtained and $P^{[4]}(x)$ will be of degree 2 at most, and will vanish at three points. The new polynomial

$$(40)\qquad y_0 + \mu\Delta_0' \frac{x}{h} \frac{1}{2!}\Delta_0'' \frac{x^2}{h^2} + \frac{1}{3!}\mu\Delta_0''' \frac{x}{h}\left(\frac{x^2}{h^2}-1\right) + \frac{1}{4!}\Delta_0^{IV}\frac{x^2}{h^2}\left(\frac{x^2}{h^2}-1\right)$$

is therefore identical with $P^{[4]}(x)$.

Since, for $i = j = n$, formulas (30) and (31) yield

$$(41)\qquad \frac{2nA_{nn}}{2D}\mu\Delta_0^{[2n-1]}\frac{x^{2n-1}}{h^{2n-1}} + \frac{A_{nn}}{2D}\Delta_0^{[2n]}\frac{x^{2n}}{h^{2n}} = \frac{1}{(2n-1)!}\mu\Delta_0^{[2n-1]}\frac{x^{2n-1}}{h^{2n-1}} + \frac{1}{(2n)!}\Delta_0^{[2n]}\frac{x^{2n}}{h^{2n}},$$

the argument by which (40) is obtained from (39) may be extended indefinitely, yielding the well known Stirling interpolation formula.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
CAMBRIDGE, MASS.

ON COVARIANTS OF LINEAR ALGEBRAS*

BY

C. C. MACDUFFEE

INTRODUCTION

This paper is concerned with covariants and concomitants of linear algebras in which neither the commutative nor the associative law of multiplication is assumed to hold. It is shown in § 2 that every algebra has associated with it a trilinear form whose covariants coincide with the totality of covariants of the algebra. In § 3 two covariants are calculated whose identical vanishing indicates, respectively, that multiplication is commutative and associative. In §§ 4, 5 the notion of identical equation of a linear algebra is generalized, and a method is developed of characterizing by covariants the properties of linear algebras with respect to these equations.

In the concluding paragraphs 6 and 7 some special cases of the above theory are worked out, and two points of contact with the existing literature are found.

CONCOMITANTS OF LINEAR ALGEBRAS

1. The definition of concomitant. Consider the linear algebra in n units $e_1, e_2, \ldots, e_n$ whose general number and multiplication table are given by, respectively,

$$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n, \tag{1}$$

$$e_i e_j = \sum_{k=1}^{k=n} c_{ijk} e_k \qquad (i, j = 1, \ldots, n). \tag{2}$$

The coefficients x_i, c_{ijk} are elements of a general field F.

We apply to the units the linear homogeneous transformation

$$e_i = \sum_{j=1}^{j=n} \alpha_{ij} e'_j, \qquad \alpha \equiv |\alpha_{ij}| \neq 0 \qquad (i = 1, \ldots, n) \tag{3}$$

where the α_{ij} are independent variables in F. Since the general number (1) is to be unaltered by transformation (3), its coefficients x_i are subject to the induced transformation

$$x'_j = \sum_{i=1}^{i=n} \alpha_{ij} x_i \qquad (j = 1, \ldots, n). \tag{4}$$

* Presented to the Society, December 28, 1922.

That is, *the coefficients x_i of the general number* (1) *are transformed by* (3) *contragrediently to the units e_i.*

We define the elements c'_{ijk} by the relation

$$e'_i\, e'_j = \sum_{k=1}^{k=n} c'_{ijk}\, e'_k \qquad (i, j = 1, \ldots, n) \tag{5}$$

and find by making use of (3) and (2) and the fact that the units are linearly independent that the c_{ijk} and c'_{ijk} are connected by the relation

$$\sum_k c_{ijk}\ \alpha_{kr} = \sum_{p,q} \alpha_{ip}\ \alpha_{jq}\ c'_{pqr}.$$

This is equivalent to

$$\alpha^2\, c'_{pqr} = \sum_{i,j,k=1}^{n} \alpha_{kr}\, A_{ip}\, A_{jq}\, c_{ijk} \quad (p, q, r = 1, \ldots, n), \tag{6}$$

where A_{ip} is the cofactor of α_{ip} in the determinant α.

We define a concomitant of the linear algebra to be a function F such that a relation

$$F(c_{ijk};\ x^{(1)}_{l_1}, x^{(2)}_{l_2}, \ldots;\ u^{(1)}_{q_1}, u^{(2)}_{q_2}, \ldots) \equiv KF(c'_{ijk};\ x^{(1)'}_{l_1}, x^{(2)'}_{l_2}, \ldots;\ u^{(1)'}_{q_1}, u^{(2)'}_{q_2}, \ldots)$$

holds identically in view of (6), the x's being cogredient with the coefficients (4) and the u's being cogredient with the units (3); K is a function of the coefficients α_{ij} of transformation. In particular if F involves only the c's it is an invariant; if it involves only the c's and x's it is a covariant; if it involves only the c's and u's it is a contravariant.

Invariantive functions which involve the units e_i have been called vector covariants by O. C. Hazlett,* who has studied many of their properties. Evidently such vector covariants are included in the above definition of concomitant.

2. **The fundamental trilinear form.** Consider the trilinear form

$$\sum_{i,j=1}^{n} x^{(1)}_i\, x^{(2)}_j\, u_k\, c_{ijk} \tag{7}$$

where the u's are independent variables cogredient with the units, and the x's are cogredient with the coefficients (4). The induced transformation deter-

* These Transactions, vol. 19 (1918), p. 408.

mined by (7) is precisely (6). We call (7) the *fundamental trilinear form** of the algebra. The following theorem is now evident.

THEOREM 1. *The totality of concomitants of any linear algebra is exactly the totality of concomitants of its fundamental trilinear form.*

Thus the problem of finding the concomitants of a linear algebra is reduced to the more familiar problem of finding the concomitants of an algebraic form. Certain known properties of concomitants of algebraic forms must be true in the case of linear algebras, as for instance the fact that K is a power of α.

3. **Concomitants which characterize commutativity and associativity.** A linear algebra is commutative provided that the units comply with the condition

$$e_i e_j = e_j e_i \qquad (i, j = 1, \ldots, n).$$

This condition on the units is readily found to be equivalent to

$$c_{ijk} - c_{jik} = 0 \qquad (i, j, k = 1, \ldots, n). \tag{8}$$

From (6) we obtain

$$c'_{ijk} - c'_{jik} = \sum_{p,q,r} \frac{A_{pi}}{\alpha} \cdot \frac{A_{qj}}{\alpha} \cdot \alpha_{rk} (c_{pqr} - c_{qpr}) \qquad (i, j, k = 1, \ldots, n).$$

The coefficients of the α's in this expression are precisely the left members of conditions (8).

By a second transformation of determinant β we have

$$c'_{ijk} - c'_{jik} = \sum_{p,q,r} \frac{B_{pi}}{\beta} \cdot \frac{B_{qj}}{\beta} \cdot \beta_{rk} (c_{pqr} - c_{qpr}).$$

That is, the equation

$$\sum_{p,q,r} \frac{A_{pi}}{\alpha} \cdot \frac{A_{qj}}{\alpha} \cdot \alpha_{rk} (c_{pqr} - c_{qpr}) = \sum_{p,q,r} \frac{B_{pi}}{\beta} \cdot \frac{B_{qj}}{\beta} \cdot \beta_{rk} (c''_{pqr} - c''_{qpr})$$

holds identically in view of the relation

$$\alpha_{ik} = \sum_j \gamma_{ij} \beta_{jk}$$

which defines the γ's, and a relation analogous to (6) which defines the c''_{ij}. Thus the α_{ij} are transformed by the γ-transformation cogrediently with the

* This form is used by Frobenius, *Theorie der hyperkomplexen Größen*, Sitzungsberichte der königlich preußischen Akademie der Wissenschaften, 1903, p. 506.

units (3) for every k and the expressions A_{ik}/α are transformed cogrediently with the coefficients (4) of the general number. Hence

$$\sum_{p,q,r} x_p^{(1)}\, x_q^{(2)}\, u_r\,(c_{pqr} - c_{qpr}) \tag{9}$$

is an absolute concomitant, and we have

THEOREM 2.* *A necessary and sufficient condition that the algebra* (2) *shall be commutative is that the concomitant* (9) *shall vanish identically.*

An algebra is associative if and only if

$$(e_i\, e_j)e_k = e_i\,(e_j\, e_k) \qquad (i, j, k = 1, \ldots, n).$$

This is equivalent to

$$\sum_{f=1}^{f=n} (c_{ijf}\, c_{fkl} - c_{jkf}\, c_{ifl}) = 0 \qquad (i, j, k, l = 1, \ldots, n). \tag{10}$$

From (6) we obtain

$$\sum_f (c'_{ijf}\, c'_{fkl} - c'_{jkf}\, c'_{ifl}) = \sum_{p,q,r,s} \frac{A_{pi}}{\alpha} \cdot \frac{A_{qj}}{\alpha} \cdot \frac{A_{rk}}{\alpha} \cdot \alpha_{sl} \sum_g (c_{pqg}\, c_{grs} - c_{qrg}\, c_{pgs})$$
$$(i, j, k, l = 1, \ldots, n).$$

As in the preceding paragraph we see that

$$\sum_{p,q,r,s} x_p^{(1)}\, x_q^{(2)}\, x_r^{(3)}\, u_s \sum_g (c_{pqg}\, c_{grs} - c_{qrg}\, c_{pgs}) \tag{11}$$

is an absolute concomitant whose coefficients vanish under conditions (10). Therefore we have

THEOREM 3.* *A necessary and sufficient condition that the algebra* (2) *shall be associative is that the concomitant* (11) *shall vanish identically.*

4. **A generalization of the notion of power series.** In an algebra in which multiplication is not always associative, a power x^n of x may be defined as a product of n equal factors x grouped according to any definite scheme. For example, there are 5 distinct fourth powers of x, namely

$$(x^3x)x, \quad x(x^3x), \quad (xx^3)x, \quad x(xx^3), \quad x^2x^2.$$

Let p_1 be any hypercomplex number. We shall define a *power series* in p_1 of order r as an ordered set $p_1, p_2, \ldots, p_r$ of powers of p_1 such that every p_i

* These results are also readily obtainable by symbolic methods.

is a product of two members (not necessarily distinct) which precede it in the series. In particular the right- and left-hand powers of x

$$x^i = x^{i-1}.x, \qquad {}^ix = x.{}^{i-1}x \qquad (i = 1, \ldots, r)$$

form power series in x.

5. **A method of deriving a class of covariants.** Consider any particular linear algebra A' in n units $e_1', e_2', \ldots, e_n'$, any polynomial p_1' and any power series $p_1', p_2', \ldots, p_r'$ where $r \leqq n$. It may or may not be possible to transform the algebra A' into an algebra A in which $p_1, p_2, \ldots, p_r$ are linearly independent. The property of being so transformable is obviously an invariantive property of the algebra A'. For if T is a transformation carrying A' into A, we may write $(A')\,T = A$. If now $(A')S = A''$ is any arbitrary transform of A', we have

$$(A'')S^{-1}T = A.$$

That is, a transformation $S^{-1}T$ exists which will carry A'' into A.

We shall now find a rational integral covariant Φ of the algebra A' whose identical vanishing is a necessary and sufficient condition that $p_1, p_2, \ldots, p_r$ be linearly independent for every transform A of A'.

If it is possible to make the transformation $(A')\,T = A$ so that $p_1, p_2, \ldots, p_r$ shall be linearly independent, it can be made in such a way that

$$p_1 = e_1, \quad p_2 = e_2, \quad \ldots, \quad p_r = e_r,$$

for every p_i is expressible as a linear function of the units. Since $p_1, p_2, \ldots, p_r$ form a power series, this implies that

$$e_i = e_j e_k \qquad (j, k < i;\ i = 2, \ldots, r).$$

From (3) and (5) we have

$$\sum_p \alpha_{ip}\, e_p' = \sum_{q,s,p} \alpha_{jq}\, \alpha_{ks}\, c_{qsp}'\, e_p'$$

and, since the units are linearly independent,

$$\alpha_{ip} = \sum_{q,s=1}^{n} \alpha_{jq}\, \alpha_{ks}\, c_{qsp}' \qquad (j, k < i;\ i = 2, \ldots, r;\ p = 1, \ldots, n). \tag{12}$$

This is a recursion formula for the α_{ip}, $i \leqq r$, in terms, ultimately, of the α_{1p}.

The determinant α of transformation T may by means of (12) be written

$$(13) \quad \alpha = \Phi(\alpha_{1i};\ \alpha_{jk};\ c'_{pqs}) \quad (i, k, p, q, s = 1, \ldots, n;\ j = r+1, \ldots, n).$$

If this polynomial is not identically zero in the field F, it is possible to choose a set of parameters α_{1i}, α_{jk}, $j > r$, in F such that $\alpha \neq 0$, and then from (12) calculate the remaining parameters α_{jk}, $j = 2, \ldots, r$, of a transformation T reducing A' to A. On the other hand, if $\Phi \equiv 0$ no such transformation is possible. Then *a necessary and sufficient condition that T exist is that* (13) *shall not vanish identically in F.*

Suppose now that the units be subjected to the transformation

$$(14) \qquad e''_i = \sum_{j=1}^{j=n} \gamma_{ij}\, e'_j \qquad (i = 1, \ldots, n)$$

of determinant $\gamma \neq 0$, the γ_{ij} being otherwise independent variables of F. The resulting transform of A' we shall call A''. The units are now e''_i and the coefficients c''_{ijk}. The situation is formally as before, and the determinant of a transformation

$$e_i = \sum_{j=1}^{j=n} \beta_{ij}\, e''_j \qquad (i = 1, \ldots, n)$$

carrying A'' into A is

$$\beta = \Phi\ (\beta_{1i};\ \beta_{jk};\ c''_{pqs}) \quad (i, k, p, q, s = 1, \ldots, n;\ j = r+1, \ldots, n).$$

Since $\alpha = \gamma\beta$, it follows that

$$(15) \qquad \Phi(\alpha_{1i};\ \alpha_{jk};\ c'_{pqs}) \equiv \gamma\, \Phi(\beta_{1i};\ \beta_{jk};\ c''_{pqs}).$$

This is an identity in view of the relation

$$\alpha_{ik} = \sum_{j=1}^{j=n} \gamma_{jk}\, \beta_{ij} \qquad (i, k = 1, \ldots, n).$$

The transformation on the x's induced by (14) is

$$x'_k = \sum_{j=1}^{j=n} \gamma_{jk}\, x''_j \qquad (k = 1, \ldots, n).$$

Let us denote by $x_k^{(i)\prime}$, for $i = 1, \ldots, n$, sets of variables cogredient with x_k'. Then we have by a mere change of notation

$$\Phi(x_i^{(1)\prime};\ x_k^{(j)\prime};\ c'_{pqs}) \equiv \gamma\Phi(x_i^{(1)\prime\prime};\ x_k^{(j)\prime\prime};\ c''_{pqs}), \tag{16}$$

which is identically true in view of the relation

$$x_k^{(i)\prime} = \sum_{j=1}^{j=n} \gamma_{jk}\, x_j^{(i)\prime\prime} \qquad (i,\ k = 1, \ldots, n).$$

Hence we have

THEOREM 4. *$\Phi(x_i^{(1)\prime};\ x_k^{(j)\prime};\ c'_{pqs})$ is a rational integral relative covariant of weight 1 in $n-r+1$ cogredient sets of variables whose identical vanishing is a necessary and sufficient condition that the series of powers $p_1, \ldots, p_r$ shall be linearly dependent for every transform A of the algebra A'.*

SOME PARTICULAR EXAMPLES

6. **The rank covariants.** It was noted in the preceding paragraph that the right- and left-hand powers of x form power series. For every linear algebra there is an equation of minimum degree in right-hand (left-hand) powers of x satisfied by the general number x of the algebra, called the right-hand (left-hand) identical equation or rank equation.* The degree of this equation is called the right-hand (left-hand) rank of the algebra, and is known to be an arithmetic invariant. By the method of § 5 of this paper we can find a covariant Φ_{ni} whose identical vanishing is a necessary and sufficient condition that the right-hand (left-hand) rank of the algebra be less than i. If now we form the sequence of covariants $(\Phi_{n,n}, \Phi_{n,n-1}, \ldots, \Phi_{n,2})$, the subscript of the first of these covariants which does not vanish identically is the right-hand (left-hand) rank of the algebra. By this means it is possible to replace the arithmetic rank invariant by an invariantive condition depending only upon the constants of multiplication c_{ijk}.

7. **The characterization of algebras of rank 2 with a principal unit.** If an algebra has a principal unit e_0 such that $e_0 x = x e_0 = x$ for every number x of the algebra, its multiplication table is given by

$$e_i\, e_j = \sum_{k=0}^{k=n} c_{ijk}\, e_k \qquad (c_{i0k} = c_{0ik} = \delta_{ik};\ i, j = 0, \ldots, n)$$

* L. E. Dickson, *Linear Algebras*, Cambridge Tract No. 16, Cambridge, 1914, p. 23.

where δ_{ik} is 0 or 1 according as $i \neq k$ or $i = k$. To such an algebra it is usually more convenient to apply, not a general linear homogeneous transformation, but a transformation of the type

$$e_i = \sum_{j=0}^{j=n} \alpha_{ij} e'_j \qquad (\alpha_{0k} = \delta_{0k};\ i = 0, \ldots, n)$$

which leaves the principal unit invariant. Evidently all the results of this paper hold for algebras having a principal unit under transformations of this type if we extend the range of the subscripts to $0, \ldots, n$ and substitute δ_{ik} for c_{i0k} and c_{0ik}, and δ_{0k} for α_{0k} throughout.

By way of illustration of our theory, we may construct a covariant Φ_{n3} whose identical vanishing indicates that every number of the algebra satisfies a quadratic equation. If the rank is $\geqq 3$, then $1, x, x^2$ are linearly independent and may be taken for the units e_0, e_1, e_2. Corresponding to (12) we have

$$\alpha_{2i} = \sum_{j,k=0}^{n} \alpha_{1j}\, \alpha_{1k}\, c'_{jki} \qquad (i = 0, \ldots, n).$$

The determinant α of the transformation may be written

$$\alpha = \sum_{i=1}^{i=n} \alpha_{2i}\, A_{2i}$$

where A_{2i} is the cofactor of α_{2i} in α. Then we may write

$$\alpha = \sum_{i=1}^{n} \sum_{j,k=0}^{n} \alpha_{1j}\, \alpha_{1k}\, A_{2i}\, c'_{jki}.$$

Since $c'_{j0i} = c'_{0ji} = \delta_{ij}$, it is readily found that

$$\alpha = \sum_{i,j,k=1}^{n} \alpha_{1i}\, \alpha_{1j}\, A_{2k}\, c'_{ijk}.$$

By substituting $x_j^{(i)}$ for α_{ij} in this expression we obtain a relative covariant of weight 1 whose identical vanishing indicates that every element of the algebra satisfies a quadratic equation.

When $n = 2$ we obtain in this way the covariant

$$\Phi_{23} = x_1^3 c_{112} + x_1^2 x_2 (c_{122} + c_{212} - c_{111}) + x_1 x_2^2 (c_{222} - c_{121} - c_{211}) - x_2^3 c_{221}$$

whose identical vanishing indicates that the rank of the algebra is less than 3. This covariant was first obtained by the author* by other methods.

For $n > 2$ we may obtain a simpler form of the concomitant Φ_{n3} by noting that the cofactors A_{2i} are transformed cogrediently, save for the factor γ^{-1}, with the units. Therefore we substitute x_i for α_{1i} and u_k for A_{2k}, and obtain the absolute concomitant

$$\Phi_{n3} = \sum_{i,j,k=1}^{n} x_i x_j u_k c_{ijk}$$

whose identical vanishing indicates that every element of the algebra satisfies a quadratic equation. For $n > 2$ the u_k are independent of the x_i and contragredient to them.

It is interesting to note that Φ_{n3} is obtainable from the fundamental trilinear form (7) by identifying the two sets of cogredient x's.

For $n > 2$ the form Φ_{n3} vanishes identically when and only when

$$c_{ijk} + c_{jik} = 0 \qquad (i, j, k = 1, \ldots, n).$$

Therefore $e_i e_j + e_j e_i = c_{ij0} + c_{ji0}$ is an element of F. This result was obtained by L. E. Dickson† for every n. He showed moreover that by another transformation we can make $e_i e_j + e_j e_i = 0$ for $i \neq j$. The present paper shows that for $n > 2$ the property that $e_i e_j + e_j e_i$ is an element of F is invariantive, while for $n = 2$ the invariantive relation indicating that every element satisfies a quadratic equation is given by the identical vanishing of Φ_{23}.

* These Transactions, vol. 23, p. 146.

† These Transactions, vol. 13, p. 63.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

A GENERALIZED PROBLEM IN WEIGHTED APPROXIMATION*

BY

DUNHAM JACKSON

1. **Introduction.** In a recent note,† the writer has discussed the convergence of the trigonometric sums $T_n(x)$ determined by the condition that

$$\int_0^{2\pi} \rho(x)[f(x)-T_n(x)]^2\,dx$$

shall be a minimum, where $f(x)$ is a given continuous function, and $\rho(x)$ is a given bounded function with a positive lower bound. The purpose of the present paper is to discuss the somewhat less simple problem of convergence that arises if the greatest lower bound of $\rho(x)$ is zero. There will be occasion incidentally to develop certain other topics, which have some independent interest: in §§ 2 and 3, the theory of the existence and uniqueness of discontinuous approximating functions of a high degree of generality, and in § 5, some corollaries of Bernstein's theorem on the derivative of a trigonometric sum.

2. **General theorem on existence of an approximating function.** Let

$$p_1(x),\ p_2(x),\ \ldots,\ p_n(x)$$

be n functions of x, defined and linearly independent for $a \leqq x \leqq b$, but otherwise completely arbitrary. It is well known, and readily proved,‡ that if

* Presented to the Society, October 27 and December 29, 1923.

† D. Jackson, *Note on the convergence of weighted trigonometric series*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 259–263. This paper will be designated by the letter A; the reader is referred to it for other bibliographical citations.

‡ Suppose the determinant were identically zero. On this assumption, let $r+1$ (equal to or less than n) be a number such that the determinants similarly formed with $r+1$ of the n functions and $r+1$ points are identically zero, for every choice of the $r+1$ functions, while it is possible to find r of the functions and r points for which the corresponding r-rowed determinant is not zero. Without loss of generality, the r functions may be denoted by $p_1, \ldots, p_r$, and the points by $x_1, \ldots, x_r$. The $(r+1)$-rowed determinant having $p_1(x_j),\ p_2(x_j),\ \ldots,\ p_{r+1}(x_j)$ for its jth row, $j = 1, 2, \ldots, r$, and $p_1(x), p_2(x), \ldots, \ldots, p_{r+1}(x)$ for its last row, is identically zero when $x_1, \ldots, x_r$ are held fast and x is

$x_1, \ldots, x_n$ are allowed to range independently over the interval, the determinant having $p_k(x_j)$ for its typical element is not identically zero. This being recognized, let $x_1, \ldots, x_n$ henceforth stand for a particular set of points for which the determinant is different from zero, the points being chosen once for all and then held fast. Let D stand for the absolute value of the determinant formed with $x_1, \ldots, x_n$, so that $D > 0$, and let D_1 be the greatest absolute value attained by any of its first minors.

Let

$$\varphi(x) = c_1 p_1(x) + c_2 p_2(x) + \cdots + c_n p_n(x)$$

be an arbitrary linear combination of the p's with constant coefficients. If H_1 is the largest of the numbers $|\varphi(x_1)|, \ldots, |\varphi(x_n)|$, and if the c's are regarded as obtainable by Cramer's rule from the simultaneous equations

$$c_1 p_1(x_j) + c_2 p_2(x_j) + \cdots + c_n p_n(x_j) = \varphi(x_j) \qquad (j = 1, 2, \ldots, n),$$

it is clear that $|c_k| \leqq n D_1 H_1 / D$ for $k = 1, 2, \ldots, n$. It is still not necessary to assume that the p's are bounded, and even if they are, $|\varphi(x)|$ may take on values larger than H_1 at other points than $x_1, \ldots, x_n$. For the applications in this paper, however, it will be sufficient to use the following form of

SIBIRANI'S LEMMA.* *If the functions $p_1, \ldots, p_n$ are linearly independent and bounded for $a \leqq x \leqq b$, and if H is an upper bound for $|\varphi(x)|$ in the interval, then*

$$|c_k| \leqq PH \qquad (k = 1, 2, \ldots, n),$$

where P is a number independent of the coefficients c_k.

---

allowed to vary over the interval; its expansion according to the elements of the last row gives a linear relation connecting $p_1(x), \ldots, p_{r+1}(x)$, in which the coefficient of p_{r+1} is not zero. This standard proof is repeated here just for the sake of assuring the reader that it does not require that the p's be continuous, bounded, measurable, or otherwise restricted in any way. The identical vanishing of the determinant is clearly a necessary as well as a sufficient condition for linear dependence, with the same complete generality.

* F. Sibirani, *Sulla rappresentazione approssimata delle funzioni*, Annali di Matematica, ser. 3, vol. 16 (1909), pp. 203–221; see p. 208. A different proof, somewhat less general in its scope, has been given by the writer: D. Jackson, *On functions of closest approximation*, these Transactions, vol. 22 (1921), pp. 117–128; see pp. 118–119. The proof given here is Sibirani's own, loc. cit., though he was not concerned with so general a statement of the conclusion. Cf. also L. Tonelli, *I polinomi d'approssimazione di Tchebychev*, the same Annali, ser. 3, vol. 15 (1908), pp. 47–119; pp. 61–62. The method is applicable also, as Sibirani makes clear, to functions of more than one variable.

A different number P might be obtained by a different choice of the points $x_1, \ldots, x_n$, but the essential thing is that such a number exists.

It will be assumed from now on that the functions p_k are bounded and measurable, and that they satisfy something more than the bare requirement of linear independence, to the extent that, for every choice of the coefficients c_k (other than $c_k = 0$, $k = 1, 2, \ldots, n$), there is a point set *of positive measure* throughout which $\varphi(x)$ is different from zero. For brevity, unless a better term is already in use, they may then be called *properly independent.*

Let H be the *least* upper bound of $|\varphi(x)|$. The functions p_k being given, H is then a function of the coefficients c_k. It will be shown that H is a *continuous* function of the c's. For the time being, the dependence of φ on the c's will be indicated explicitly by the notation $\varphi(x, c_1, \ldots, c_n)$. Let Q be a number such that $|p_k(x)| \leqq Q$ for $a \leqq x \leqq b$ and for $k = 1, 2, \ldots, n$. If c is the largest of the numbers $|c_k|$,

$$|\varphi(x, c_1, \ldots, c_n)| \leqq n c Q .$$

Let ε be an arbitrary positive quantity. Let two sets of coefficients be considered, (c_k) and $(c_k + \Delta c_k)$, and suppose that $|\Delta c_k| \leqq \varepsilon/(2nQ)$ for $k = 1, 2, \ldots, n$. Then, since φ depends linearly on the c's,

$$|\varphi(x, c_1 + \Delta c_1, \ldots, c_n + \Delta c_n) - \varphi(x, c_1, \ldots, c_n)| = |\varphi(x, \Delta c_1, \ldots, \Delta c_n)| < \frac{\varepsilon}{2}$$

throughout (a, b). As $|\varphi(x, c_1, \ldots, c_n)| \leqq H(c_1, \ldots, c_n)$ throughout the interval, it follows that

$$|\varphi(x, c_1 + \Delta c_1, \ldots, c_n + \Delta c_n)| < H(c_1, \ldots, c_n) + \frac{\varepsilon}{2} ;$$

that is,

$$H(c_1 + \Delta c_1, \ldots, c_n + \Delta c_n) < H(c_1, \ldots, c_n) + \frac{\varepsilon}{2} .$$

On the other hand, since H is the *least* upper bound of φ, there will be a point x_0 where

$$|\varphi(x_0, c_1, \ldots, c_n)| > H(c_1, \ldots, c_n) - \frac{\varepsilon}{2} .$$

Then $|\varphi(x_0, c_1 + \Delta c_1, \ldots, c_n + \Delta c_n)| > H(c_1, \ldots, c_n) - \varepsilon$, and

$$H(c_1 + \Delta c_1, \ldots, c_n + \Delta c_n) > H(c_1, \ldots, c_n) - \varepsilon .$$

This proves the continuity of H as a function of the c's.

It will be shown next that if J denotes the integral

$$J = \int_a^b |\varphi(x)|^m \, dx,$$

for a fixed $m > 0$, and if the coefficients in φ are allowed to vary in such a way that J remains bounded, the coefficients themselves will remain bounded also. Sibirani's lemma establishes the fact that the sets of coefficients for which the least upper bound H has a given fixed value, $H = 1$, say, correspond to a bounded point set in the space of n dimensions having $c_1, \ldots, c_n$ for coördinates. It follows from the last paragraph that this point set is closed. On the other hand, J itself is a continuous function of the coefficients. Among the functions φ for which $H = 1$, then, there will be at least one for which the value of J is a minimum. Let this minimum be denoted by A. It is certain that $A > 0$; the coefficients in φ can not all be zero when $H = 1$, and therefore, since the p's are assumed properly independent, φ is different from zero throughout a set of positive measure. If φ is an arbitrary linear combination of the p's (not identically zero), and H the least upper bound of its absolute value, the least upper bound of $|\varphi/H|$ is 1, and

$$\int_a^b \left| \frac{\varphi(x)}{H} \right|^m dx \geqq A, \qquad J = \int_a^b |\varphi(x)|^m \, dx \geqq AH^m.$$

Let c once more be equal to the largest $|c_k|$; then

$$c \leqq PH, \qquad H \geqq \frac{c}{P}, \qquad J \geqq \frac{Ac^m}{P^m}, \qquad c \leqq P\left(\frac{J}{A}\right)^{1/m},$$

or, if P_1 stands for the number $P/A^{1/m}$, independent of the c's,

$$c \leqq P_1 J^{1/m}.$$

This proves that the c's are bounded when J is bounded.

Let $f(x)$ be an arbitrary function, bounded and measurable in (a, b). It is possible now to consider the minimizing of the integral

$$J' = \int_a^b |f(x) - \varphi(x)|^m \, dx.$$

This integral is a continuous function of the coefficients in φ. Suppose first that f and the p's constitute a set of $n+1$ properly independent functions. If J_0' is a particular value taken on by J', the coefficients for which $J' \leqq J_0'$ correspond to a point set which is bounded, by the preceding paragraph, and closed, because of the continuity of J'. There is consequently a determination of the coefficients for which J' is a minimum. If f and the p's are not properly independent, a linear relation connecting them must involve f with a non-vanishing coefficient, because of the independence of the p's by themselves. This means that f is identically equal to a linear combination of the p's in (a, b), except possibly for a point set of measure zero. If this linear combination of the p's is taken for φ, J' is reduced to zero, which is then clearly its minimum value. In summary:

If the functions $p_1(x), \ldots, p_n(x)$ are bounded, measurable, and properly independent for $a \leqq x \leqq b$, if $f(x)$ is bounded and measurable in the same interval, and if m is a positive constant, there is at least one determination of the coefficients in the linear combination $\varphi = c_1 p_1 + \cdots + c_n p_n$ for which the value of the integral

$$\int_a^b |f(x) - \varphi(x)|^m \, dx$$

*is a minimum.**

A function φ which gives the integral its minimum value will be called an *approximating function.*

3. **General theorem on uniqueness of the approximating function, $m > 1$.** A familiar type of argument will justify the assertion

If $m > 1$ in the hypothesis of the preceding theorem, the approximating function is uniquely determined.

Suppose, if possible, that φ_1 and φ_2 are two approximating functions which are not identically equal. Let the dependence of J' on φ be indicated by the notation $J'(\varphi)$, and let the minimum value of J' be denoted by γ. Then

$$J'(\varphi_1) = J'(\varphi_2) = \gamma.$$

* This theorem would still be true if the p's were not independent. The truth of the statement is trivial if they are all identically zero. Otherwise, there is a properly independent subset among them, on which the remaining functions of the set are linearly dependent, except possibly for point sets of measure zero. There is an approximating function in terms of the properly independent subset, by the theorem as proved above, and this is at the same time an approximating function in terms of the original set of p's. But the hypothesis of independence is necessary for the theorem of uniqueness in the next section.

Let $\varphi_3 = \frac{1}{2}(\varphi_1 + \varphi_2)$, so that $f - \varphi_3 = \frac{1}{2}[(f-\varphi_1)+(f-\varphi_2)]$. In consequence of the property which is described geometrically by saying that the graph of the function $Y = |X|^m$ is everywhere concave upwards (when $m > 1$),

$$|f-\varphi_3|^m \leqq \frac{1}{2}[|f-\varphi_1|^m + |f-\varphi_2|^m] \tag{1}$$

for all values of x, and the relation is an actual inequality at every point where $\varphi_1 \neq \varphi_2$. If φ_1 and φ_2 are not identical, they differ in value over a point set of positive measure, by the hypothesis of proper independence of the p's, and the sign of inequality holds in (1) at all points of this set. Consequently

$$J(\varphi_3) < \frac{1}{2}[J(\varphi_1) + J(\varphi_2)];$$

that is, $J(\varphi_3) < \gamma$, which is impossible.

4. Existence and uniqueness of solution of a problem in weighted trigonometric approximation. The general theorems above have been derived here primarily for the sake of a more specific application. Let $\varrho(x)$ be a function which is bounded and measurable, and of period 2π; let it be assumed that $\varrho(x) \geqq 0$ everywhere, and that $\varrho(x) > 0$ over a point set of positive measure in an interval of length 2π. Let $f(x)$ be a function which is bounded and measurable, and of period 2π, but otherwise unrestricted, for the present. Let $T_n(x)$ be a trigonometric sum of the nth order,* and let m be a positive constant. The discussion will be concerned now with the integral

$$\int_0^{2\pi} \varrho(x)\,|f(x) - T_n(x)|^m\,dx. \tag{2}$$

The most interesting case, mentioned in the introduction, is that in which $m = 2$, but the method applies equally well to the more general problem.

The integral (2) can be written in the form

$$\int_0^{2\pi} |\sqrt[m]{\varrho(x)}\,f(x) - \sqrt[m]{\varrho(x)}\;T_n(x)|^m\,dx.$$

* This will be understood throughout to mean a trigonometric sum of order n *at most*.

The function $\sqrt[m]{\varrho(x)}f(x)$ is bounded and measurable for $0 \leqq x \leqq 2\pi$. The expression $\sqrt[m]{\varrho(x)}\ T_n(x)$ is a linear combination of the functions

$$(3) \qquad \sqrt[m]{\varrho(x)}, \qquad \sqrt[m]{\varrho(x)}\cos kx, \qquad \sqrt[m]{\varrho(x)}\sin kx \qquad (k = 1, 2, \ldots . n),$$

and these functions are bounded, measurable, and properly independent for $0 \leqq x \leqq 2\pi$, since, if $T_n(x)$ is any trigonometric sum of the nth order with coefficients not all zero, $\sqrt[m]{\varrho(x)}\ T_n(x)$ is different from zero at all points of the set where $\varrho \neq 0$, with the exception of $2n$ points at most in any interval of length 2π. So the preceding theorems are applicable, with n replaced by $2n+1$, the functions $p_k(x)$ by the functions (3), and $f(x)$ by $\sqrt[m]{\varrho(x)}f(x)$, while $a = 0$, $b = 2\pi$. The conclusion is as follows:*

Under the hypotheses stated at the beginning of the section, there will be at least one, and, when $m > 1$, just one determination of the coefficients in $T_n(x)$, for which the value of the integral (2) *is a minimum.*†

5. **Corollaries of Bernstein's Theorem.** The convergence proof in the paper A depended essentially on

BERNSTEIN'S THEOREM.‡ *If $T_n(x)$ is a trigonometric sum of order n, and if $|T_n(x)| \leqq L$ for all values of x, then $|T_n'(x)|$ can never exceed nL.*

This theorem is not immediately adequate in the present circumstances, inasmuch as it requires that the condition on $|T_n(x)|$ be satisfied for *all* values of x, or, what amounts to the same thing, throughout an interval of length 2π, if the conclusion is to be available over even a part of a period; and there will be occasion now to reason from a similar hypothesis, the validity of which

* Mr. J. Shohat informs me that he had obtained an analogous result some years ago, in an unpublished paper, for the case of polynomial approximation. The present method is of course applicable to the polynomial case also.

† When $m < 1$, the determination is not necessarily unique; cf. D. Jackson, *Note on an ambiguous case of approximation*, these Transactions, vol. 25 (1923), pp. 333–337. This paper will be cited by the letter B. It can be regarded as dealing with a special case under the present discussion, in which ρ is identically 1. In this special case, the determination is unique when $m = 1$; cf. D. Jackson, *Note on a class of polynomials of approximation*, these Transactions, vol. 22 (1921), pp. 320–326. The question of uniqueness for $m = 1$ and for general ρ will not be treated here.

‡ Cf., e. g., D. Jackson, *The general theory of approximation by polynomials and trigonometric sums*, Bulletin of the American Mathematical Society, vol. 27 (1921), pp. 415–431; see pp. 421–422. The proof given in the Bulletin was quoted there as due originally to de la Vallée Poussin; it has since come to my attention that it had been published still earlier by Marcel Riesz, *Eine trigonometrische Interpolationsformel und einige Ungleichungen für Polynome*, Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 23 (1914), pp. 354–368; see pp. 360–361.

is granted only over an interval of length less than 2π. The relations that are needed can be derived indirectly from Bernstein's theorem, and it is the purpose of this section to supply them.

Let $P_n(x)$ be a polynomial of the nth degree, and L the maximum of $|P_n(x)|$ for $-1 \leqq x \leqq 1$. Let $x = \cos\theta$. Then $P_n(x) = T_n(\theta)$ is a trigonometric sum of the nth order in θ (involving only cosines), and, by application of the theorem just quoted,*

$$|\sqrt{1-x^2}\,P_n'(x)| = |\sin\theta\,P_n'(\cos\theta)| = |T_n'(\theta)| \leqq nL,$$

$$|P_n'(x)| \leqq \frac{nL}{\sqrt{1-x^2}}$$

for $-1 < x < 1$.

More generally, let $P_n(x)$ be a polynomial of the nth degree, and let L be the maximum of $|P_n(x)|$ over an arbitrary interval $a \leqq x \leqq b$. Let $y = (2x-a-b)/(b-a)$. Then $P_n(x) = Q_n(y)$ is a polynomial of the nth degree in y, and L is the maximum of $|Q_n(y)|$ for $-1 \leqq y \leqq 1$, so that $|Q_n'(y)| \leqq nL/\sqrt{1-y^2}$ throughout the interior of this interval. But $Q_n'(y) = \frac{1}{2}(b-a)P_n'(x)$, the accent denoting differentiation with respect to the indicated argument in each case; and, on the other hand,

$$1+y = \frac{2(x-a)}{b-a},\quad 1-y = \frac{2(b-x)}{b-a},\quad \frac{1}{\sqrt{1-y^2}} = \frac{b-a}{2\sqrt{(b-x)(x-a)}}.$$

Consequently

$$|P_n'(x)| \leqq \frac{nL}{\sqrt{(b-x)(x-a)}}$$

for $a < x < b$.

Now let $P_{n-1}(x)$ be a polynomial of degree $n-1$, and let L be the maximum of $|\sqrt{1-x^2}\,P_{n-1}(x)|$ for $-1 \leqq x \leqq 1$. Let $x = \cos\theta$. Then $P_{n-1}(x)$ is a cosine sum of order $n-1$ in θ. The result of multiplying this sum by $\sin\theta$ is a trigonometric sum of order n (involving sines only), which may be denoted by $T_n(\theta)$. In terms of the original variables,

* Cf. M. Riesz, loc. cit., pp. 359–360; D. Jackson, *On the convergence of certain trigonometric and polynomial approximations* (to be cited below by the letter C), these Transactions, vol. 22 (1921), pp. 158–166, see p. 162; and, originally, S. Bernstein, *Sur l'ordre de la meilleure approximation des fonctions continues par des polynomes de degré donné*, Mémoire couronné, Brussels, 1912, pp. 6–11. My paper, written before I knew of the existence of that of Riesz, should have contained a reference to the latter in connection with pp. 162–163.

$T_n(\theta) = \pm\sqrt{1-x^2}\,P_{n-1}(x)$, so that $|T_n(\theta)| \leqq L$, and $|T_n'(\theta)| \leqq nL$, for all values of θ. Inasmuch as

$$|T_n'(\theta)| = \left|\sin\theta\frac{d}{dx}T_n(\theta)\right| = \left|\sqrt{1-x^2}\,\frac{d}{dx}[\sqrt{1-x^2}\,P_{n-1}(x)]\right|,$$

it is seen that*

$$\left|\frac{d}{dx}[\sqrt{1-x^2}\,P_{n-1}(x)]\right| \leqq \frac{nL}{\sqrt{1-x^2}}$$

for $-1<x<1$.

Let $P_{n-1}(x)$ be a polynomial of degree $n-1$ again, but let L this time be the maximum of $|\sqrt{(b-x)(x-a)}\,P_{n-1}(x)|$ for $a \leqq x \leqq b$. If the change of variable of the second paragraph preceding is carried through, it is found that the various factors $\frac{1}{2}(b-a)$ cancel each other, and the conclusion is that

$$\left|\frac{d}{dx}[\sqrt{(b-x)(x-a)}\,P_{n-1}(x)]\right| \leqq \frac{nL}{\sqrt{(b-x)(x-a)}}$$

for $a<x<b$.

These lemmas with regard to polynomials have been set down here for the sake of their application to trigonometric sums. Let $U_n(\theta)$ be a trigonometric sum of the nth order involving only cosines, and let L be the maximum of $|U_n(\theta)|$ for $-\alpha \leqq \theta \leqq \alpha$, where $0<\alpha<\pi$. As $U_n(\theta)$ is a polynomial of the nth degree in $\cos\theta$, and L the maximum of its absolute value for $\cos\alpha \leqq \cos\theta \leqq 1$, one of the preceding results is applicable, to the effect that

$$\left|\frac{d}{d\cos\theta}U_n(\theta)\right| \leqq \frac{nL}{\sqrt{(1-\cos\theta)(\cos\theta-\cos\alpha)}},$$

$$|U_n'(\theta)| \leqq \frac{nL|\sin\theta|}{\sqrt{(1-\cos\theta)(\cos\theta-\cos\alpha)}} = nL\sqrt{\frac{1+\cos\theta}{\cos\theta-\cos\alpha}} \leqq \frac{\sqrt{2}\,nL}{\sqrt{\cos\theta-\cos\alpha}}.$$

The reasoning is valid in the first instance for $\cos\alpha<\cos\theta<1$; but as $U_n'(\theta)$ and the last member on the right are continuous for $\theta = 0$, the relation between them remains in force at this point also, and so holds throughout the interval $-\alpha<\theta<\alpha$.

* Cf. Bernstein, op. cit., pp. 17–19.

The corresponding argument for a sum of sines appears to be materially less simple. Let $V_n(\theta)$ be a trigonometric sum of the nth order involving only sines, and let L be the maximum of $|V_n(\theta)|$ for $-\alpha \leqq \theta \leqq \alpha$, where $0 < \alpha < \pi$, as before. The quotient $V_n(\theta)/\sin\theta$ can be expressed as a polynomial of degree $n-1$ in $\cos\theta$, and may be represented by $P_{n-1}(\cos\theta)$. Let α_1 and α_2 be two numbers between 0 and α, so that $0 < \alpha_2 < \alpha_1 < \alpha$, and so that $\alpha_2 < \pi - \alpha$. For $\cos\alpha \leqq \cos\theta \leqq \cos\alpha_2$,

$$|P_{n-1}(\cos\theta)| = \left|\frac{V_n(\theta)}{\sin\theta}\right| \leqq \frac{L}{\sin\alpha_2},$$

and, throughout the interior of the interval,

$$\left|\frac{d}{d\theta}P_{n-1}(\cos\theta)\right| = |\sin\theta\, P'_{n-1}(\cos\theta)| \leqq |P'_{n-1}(\cos\theta)|$$

$$\leqq \frac{(n-1)L}{\sin\alpha_2\sqrt{(\cos\alpha_2-\cos\theta)(\cos\theta-\cos\alpha)}}.$$

If θ is further restricted so that $\cos\alpha < \cos\theta \leqq \cos\alpha_1$, then $\cos\alpha_2 - \cos\theta \geqq \cos\alpha_2 - \cos\alpha_1$, and it can be stated briefly that

$$\left|\frac{d}{d\theta}P_{n-1}(\cos\theta)\right| \leqq \frac{k_1 n L}{\sqrt{\cos\theta-\cos\alpha}},$$

where k_1 is a number independent of n, and depending only on α_1 and α_2. In the same interval,

$$|V'_n(\theta)| = \left|\frac{d}{d\theta}[\sin\theta\, P_{n-1}(\cos\theta)]\right|$$

$$= \left|\sin\theta\frac{d}{d\theta}P_{n-1}(\cos\theta) + \cos\theta\, P_{n-1}(\cos\theta)\right| \leqq \frac{k_1 n L}{\sqrt{\cos\theta-\cos\alpha}} + \frac{L}{\sin\alpha_2}.$$

If the second term in the last member is merged with the first term by means of the trivial relations

$$1 \leqq n, \quad \frac{1}{\sin\alpha_2} = \frac{\sqrt{\cos\theta-\cos\alpha}}{\sin\alpha_2}\cdot\frac{1}{\sqrt{\cos\theta-\cos\alpha}} \leqq \frac{\sqrt{2}}{\sin\alpha_2}\cdot\frac{1}{\sqrt{\cos\theta-\cos\alpha}},$$

it is seen that

$$|V_n'(\theta)| \leqq \frac{k_2 n L}{\sqrt{\cos\theta - \cos\alpha}}$$

for $\cos\alpha < \cos\theta \leqq \cos\alpha_1$, the number k_2 depending once more on α_1 and α_2 and on nothing else.

For $\cos\alpha \leqq \cos\theta \leqq 1$,

$$|\sin\theta\, P_{n-1}(\cos\theta)| \leqq L.$$

But in this interval

$$|\sin\theta| = \sqrt{1-\cos^2\theta} = \sqrt{(1-\cos\theta)(\cos\theta-(-1))}$$

$$> \sqrt{(1-\cos\theta)(\cos\theta-\cos\alpha)},$$

and therefore

$$\left|\sqrt{(1-\cos\theta)(\cos\theta-\cos\alpha)}\, P_{n-1}(\cos\theta)\right| \leqq L.$$

Let the quantity between the bars in the last relation be denoted by $W(\theta)$. By the last of the lemmas for polynomials,

$$\left|\frac{d}{d\cos\theta} W(\theta)\right| \leqq \frac{nL}{\sqrt{(1-\cos\theta)(\cos\theta-\cos\alpha)}}$$

for $\cos\alpha < \cos\theta < 1$, whence

$$|W'(\theta)| \leqq \frac{nL|\sin\theta|}{\sqrt{(1-\cos\theta)(\cos\theta-\cos\alpha)}} \leqq \frac{\sqrt{2}\, nL}{\sqrt{\cos\theta-\cos\alpha}}.$$

It will be remembered that

$$V_n(\theta) = \sin\theta\, P_{n-1}(\cos\theta) = \pm\sqrt{\frac{1+\cos\theta}{\cos\theta-\cos\alpha}}\, W(\theta).$$

The radical, which may be denoted by $Z(\theta)$, becomes infinite for $\theta = \alpha$. If θ is restricted so that $\cos\alpha_1 \leqq \cos\theta < 1$, however, both $Z(\theta)$ and its derivative will be bounded, and it can be asserted that

$$|V_n'(\theta)| = |Z(\theta)\,W'(\theta) + Z'(\theta)\,W(\theta)|$$

$$\leqq \frac{k_3\,n\,L}{\sqrt{\cos\theta - \cos\alpha}} + k_4 L \leqq \frac{k_5\,n\,L}{\sqrt{\cos\theta - \cos\alpha}},$$

the numbers k_3, k_4, and k_5 depending only on α_1. As far as the interval now under discussion is concerned, the denominator could of course be omitted, with suitable modification of the constants; it is retained for convenience in passing on to the main conclusion. The relation between the first and last members holds by continuity for $\theta = 0$ as well.

Let k_6 be the larger of the numbers k_2 and k_5. As α_1 and α_2 can be chosen once for all when α is given, it may be considered that the k's, and in particular k_6, depend only on α. The two preceding paragraphs, taken together, show that

$$|V_n'(\theta)| \leqq \frac{k_6\,n\,L}{\sqrt{\cos\theta - \cos\alpha}}$$

for $-\alpha < \theta < \alpha$.

Finally, let $T_n(\theta)$ be an arbitrary trigonometric sum of order n. Let L be the maximum of $|T_n(\theta)|$ for $-\alpha \leqq \theta \leqq \alpha$, $0 < \alpha < \pi$. Let

$$U_n(\theta) = \frac{1}{2}[T_n(\theta) + T_n(-\theta)], \qquad V_n(\theta) = \frac{1}{2}[T_n(\theta) - T_n(-\theta)],$$

so that $U_n(\theta)$ is a sum of cosines, $V_n(\theta)$ a sum of sines, $T_n(\theta) = U_n(\theta) + V_n(\theta)$, and $|T_n'(\theta)| \leqq |U_n'(\theta)| + |V_n'(\theta)|$. It is seen that $|U_n(\theta)| \leqq L$, $|V_n(\theta)| \leqq L$, throughout the interval. The inequalities that have been obtained for $|U_n'(\theta)|$ and $|V_n'(\theta)|$ separately can be combined; if the number $\sqrt{2} + k_6$, brought in by the combination, is denoted by k_0, the conclusion is that

$$|T_n'(\theta)| \leqq \frac{k_0\,n\,L}{\sqrt{\cos\theta - \cos\alpha}}$$

for $-\alpha < \theta < \alpha$, the value of k_0 depending only on α, not on θ, or n, or the coefficients in $T_n(\theta)$.

The essence of this conclusion can also be expressed in a slightly different form, by the use of the relation

$$\cos\theta - \cos\alpha = 2\sin\frac{\alpha - \theta}{2}\sin\frac{\alpha + \theta}{2}.$$

Since $0 < \alpha \pm \theta < 2\alpha < 2\pi$ throughout the interval in question,

$$\frac{\sin \frac{1}{2}(\alpha \pm \theta)}{\frac{1}{2}(\alpha \pm \theta)} > \frac{\sin \alpha}{\alpha} > 0,$$

and

$$\cos\theta - \cos\alpha > \frac{\sin^2\alpha}{2\alpha^2}(\alpha^2 - \theta^2).$$

If $k_0 \alpha \sqrt{2}/\sin\alpha$ is denoted by k, this means that

$$|T_n'(\theta)| \leqq \frac{knL}{\sqrt{\alpha^2 - \theta^2}}.$$

The last inequality can be adapted immediately, by a linear transformation, to an interval not symmetric with respect to the origin. With a slight change of notation, the result may be summarized in the following

LEMMA. *If $T_n(x)$ is a trigonometric sum of order n, and if $|T_n(x)| \leqq L$ for $\alpha \leqq x \leqq \beta$, where $0 < \beta - \alpha < 2\pi$, then*

$$|T_n'(x)| \leqq \frac{knL}{\sqrt{(\beta - x)(x - \alpha)}}$$

for $\alpha < x < \beta$, the multiplier k depending only on the length of the interval, not on x, or n, or the coefficients in $T_n(x)$.

6. **Convergence.** With the notation and results of § 4, and with the lemma of § 5, some progress can be made in the study of the convergence of $T_n(x)$ toward $f(x)$ as n becomes infinite. There will be occasion naturally to impose further restrictions on $f(x)$, which will be specified later. It will clearly be necessary also to introduce additional hypotheses either with regard to $\varrho(x)$ or with regard to the values of x for which the convergence is to be established; for if $\varrho(x)$ were to vanish identically throughout an interval, convergence to the value of $f(x)$ could not be proved at interior points of this interval, since $f(x)$ could be replaced by an altogether different function in the interval, without affecting the determination of $T_n(x)$, and the change could be made, in the interior of the interval at least, without violating the requirements of continuity* that are to be placed on $f(x)$. It will be assumed

* It will not be assumed anywhere in this paper that $f(x)$ is analytic.

for convenience of expression that $m > 1$, so that the approximating function is known to be uniquely determined, though the method is applicable in other cases also*.

As in other discussions of similar character, the problem can be simplified a little at the outset. Let $\varphi(x)$ be a function such that the difference between f and φ is identically equal to a trigonometric sum $t_n(x)$ of the nth order:

$$\varphi(x) = f(x) - t_n(x). \tag{4}$$

If $T_n(x)$ is for the moment an arbitrary trigonometric sum of the nth order, and if

$$\tau_n(x) = T_n(x) - t_n(x), \tag{5}$$

the value of the integral (2) is the same as that of the corresponding integral with $f(x)$ and $T_n(x)$ replaced by $\varphi(x)$ and $\tau_n(x)$ respectively. In particular, if $T_n(x)$ is the approximating sum for $f(x)$—as will be understood henceforth—then $\tau_n(x)$ is at the same time the approximating sum for $\varphi(x)$. In other words, the approximating sums for $f(x)$ and $\varphi(x)$ are connected by the relation (5), and the errors $f(x) - T_n(x)$ and $\varphi(x) - \tau_n(x)$ are identical. It will be possible to choose φ so that the latter error is more convenient to deal with directly.

By way of preliminary discussion, let $\varphi(x)$ be an arbitrary bounded and measurable function of period 2π, and $\tau_n(x)$ the approximating sum of the nth order for $\varphi(x)$, $n \geqq 1$. Let ϵ be the least upper bound of $|\varphi(x)|$. For the present, ϵ may be large or small, although, as the notation suggests, a succession of functions φ will ultimately be considered, with values of ϵ approaching zero. *Let it be supposed that there is an interval* $\alpha \leqq x \leqq \beta$, *throughout which (except possibly for a point set of measure zero)* $\varrho(x) \geqq v > 0$. Let $\varrho(x) \leqq V$ for all values† of x. Let μ be the maximum of $|\tau_n(x)|$ *in the interval* $\alpha \leqq x \leqq \beta$, and let x_0 be a point of this interval such that $|\tau_n(x_0)| = \mu$. It is to be shown that a relation of inequality connecting μ and ϵ can be deduced without further specification of the function φ.

By the lemma of the preceding section,

$$|\tau_n'(x)| \leqq \frac{kn\mu}{\sqrt{(\beta - x)(x - \alpha)}} \tag{6}$$

* Cf. the paper B, § 4.

† Here again, exception may by made of a point set of measure zero, without affecting the values of the integrals involved, and so without affecting the essential conditions of the problem.

for $\alpha < x < \beta$, the multiplier k depending only on $\beta - \alpha$. Consequently, if x is a point of the interval,

$$|\tau_n(x) - \tau_n(x_0)| \leqq kn\mu \left| \int_{x_0}^{x} \frac{dx}{\sqrt{(\beta - x)(x - \alpha)}} \right|.$$

For definiteness, suppose that $\alpha \leqq x_0 \leqq \frac{1}{2}(\alpha + \beta)$; the possibility that x_0 belongs to the other half of the interval is to be treated in essentially the same way. Let h be a positive number $< \frac{1}{2}(\beta - \alpha)$. Then the interval $(x_0, x_0 + h)$ is wholly contained in (α, β). For $x_0 \leqq x \leqq x_0 + h$,

$$|\tau_n(x) - \tau_n(x_0)| \leqq kn\mu \int_{x_0}^{x_0+h} \frac{dx}{\sqrt{(\beta - x)(x - \alpha)}}.$$

The integrand is the reciprocal of the ordinate of a semicircle with its extremities on the x-axis at the points with abscissas α and β, so that the value of the integral from x_0 to $x_0 + h$ is not greater than the value of the integral of the same integrand from α to $\alpha + h$; it would of course be easy, but perhaps superfluous, to give the details of an analytical proof. Furthermore, $\beta - x > \frac{1}{2}(\beta - \alpha)$ when x is between α and $\alpha + h$, so that, in formulas,

$$\int_{x_0}^{x_0+h} \frac{dx}{\sqrt{(\beta - x)(x - \alpha)}} \leqq \int_{\alpha}^{\alpha+h} \frac{dx}{\sqrt{(\beta - x)(x - \alpha)}}$$

$$\leqq \sqrt{\frac{2}{\beta - \alpha}} \int_{\alpha}^{\alpha+h} \frac{dx}{\sqrt{x - \alpha}} = 2\sqrt{\frac{2h}{\beta - \alpha}},$$

$$|\tau_n(x) - \tau_n(x_0)| \leqq k' n\mu \sqrt{h},$$

where k' is a new multiplier depending only on $\beta - \alpha$.

Now let* $h = 1/(4k'^2 n^2)$. This h will satisfy the requirement of being less than $\frac{1}{2}(\beta - \alpha)$, as soon as n is sufficiently large; more precisely, this will be the case for $n \geqq n_0$, where n_0, like the k's, *depends only on* $\beta - \alpha$, not on ϱ or φ

* It is clear that $k' > 0$; for k, in the lemma of § 5, is certainly not zero, and k' is equal to k multiplied by a positive quantity.

or anything else. As the ultimate question at issue is one of convergence for $n = \infty$, it may be assumed that $n \geqq n_0$ henceforth. Then, for $x_0 \leqq x \leqq x_0 + h$,

$$|\tau_n(x) - \tau_n(x_0)| \leq \frac{\mu}{2}, \qquad |\tau_n(x)| \geq \frac{\mu}{2}. \tag{7}$$

This is on the hypothesis that $x_0 \leqq \frac{1}{2}(\alpha + \beta)$; a similar argument applies if $x_0 > \frac{1}{2}(\alpha + \beta)$, the conclusion holding for $x_0 - h \leq x \leq x_0$, with the same values of k', h, and n_0 as before.

Let

$$\gamma_n = \int_0^{2\pi} \varrho(x) |\varphi(x) - \tau_n(x)|^m \, dx.$$

By the definition of ϵ, $|\varphi(x)| \leqq \epsilon$ for all values of x. Let it be supposed for the moment that $\mu \geqq 4\epsilon$; the contrary hypothesis will be considered later. Then $|\varphi(x)| \leqq \mu/4$, and the relations (7) imply that

$$|\varphi(x) - \tau_n(x)| \geqq \frac{\mu}{4}.$$

These relations hold at least throughout an interval of length h, where h has the value specified above; the interval of length h is contained in the interval (α, β), where $\varrho(x) \geqq v > 0$, except possibly for a set of measure zero; and the integrand in the expression* for γ_n is never negative. So it can be inferred that

$$\gamma_n \geqq hv \left(\frac{\mu}{4}\right)^m = \frac{v\mu^m}{4^{m+1} k'^2 n^2}.$$

As $\tau_n(x)$ is the approximating sum for $\varphi(x)$, the value of the integral would not be diminished if $\tau_n(x)$ were replaced by any other trigonometric sum of order n — in particular, by zero. That is, since $|\varphi(x)| \leqq \epsilon$ and $0 \leqq \varphi(x) \leqq V$,

$$\gamma_n \leqq 2\pi V \epsilon^m.$$

* It is immaterial whether the interval (α, β) is contained in $(0, 2\pi)$ or not, as the integral defining γ_n can equally well be extended over any other interval of length 2π, because of the periodicity of the integrand. It may be assumed that $\beta - \alpha < 2\pi$, since otherwise the reasoning of the paper A would be directly applicable.

Consequently, for $n \geqq n_0$,

$$\frac{v\mu^m}{4^{m+1}k'^2 n^2} \leqq 2\pi V \varepsilon^m, \qquad \mu \leqq \sqrt[m]{\frac{8k'^2\pi V}{v}} \cdot 4\varepsilon n^{2/m} = C'\varepsilon n^{2/m},$$

where C' depends on m, V, v, and $\beta - \alpha$, but not on n or on $\varphi(x)$.

This is on the hypothesis that $\mu \geqq 4\varepsilon$. But, in the contrary case, $\mu < 4\varepsilon \leqq 4\varepsilon n^{2/m}$, since it is understood that $n \geqq 1$, and if C'' stands for the larger of the numbers 4. C', then

$$\mu \leqq C''\varepsilon n^{2/m} \tag{8}$$

without exception, as soon as $n \geqq n_0$.

It will be recalled that μ is the maximum of $|\tau_n(x)|$ for $\alpha \leqq x \leqq \beta$, and that $|\varphi(x)| \leqq \varepsilon$ for all values of x. Hence

$$|\varphi(x) - \tau_n(x)| \leqq \varepsilon + C''\varepsilon n^{2/m} \leqq \varepsilon n^{2/m} + C''\varepsilon n^{2/m},$$

or, if $C'' + 1 = C$,

$$|\varphi(x) - \tau_n(x)| \leqq C\varepsilon n^{2/m}$$

for $\alpha \leqq x \leqq \beta$ and for $n \geqq n_0$, the multiplier C, like C' and C'', being independent of x, n and φ.

To apply this result, let $\varphi(x)$ be connected with the given function $f(x)$ by an identity of the form (4). Let such a function φ be constructed for each value of $n \geqq 1$, the respective φ's being represented by the notation $\varphi_n(x)$, and the least upper bound of $|\varphi_n(x)|$ by ε_n. Let the approximating sum $\tau_n(x)$, of corresponding order n, be formed for each φ_n. As $\varphi_n(x) - \tau_n(x)$ is identical with $f(x) - T_n(x)$ for each value of n, where $T_n(x)$ is the approximating sum of the nth order for $f(x)$, it is seen that*

$$|f(x) - T_n(x)| \leqq C n^{2/m} \varepsilon_n \tag{9}$$

for $\alpha \leqq x \leqq \beta$ and for $n \geqq n_0$.

* The reader may be assailed at first by a suspicion that this argument plays fast and loose with the subscript n. The fact is, however, that for independent n and ν

$$|\varphi_n(x) - \tau_\nu(x)| \leqq C\nu^{2/m}\varepsilon_n$$

when $\nu \geqq n_0$, if $\tau_\nu(x)$ is the approximating sum of order ν for the function $\varphi_n(x)$; and then use is made of this relation, in the case of each $n \geqq n_0$, only for the single value $\nu = n$. The essential point is of course the fact that C is independent of φ.

If $f(x)$ is continuous, as must be assumed from now on, it follows from Weierstrass's theorem that the trigonometric sums $t_n(x)$ which define the remainders φ_n can be chosen so that $\lim_{n=\infty} \varepsilon_n = 0$. The rapidity with which ε_n can be made to approach zero depends on further properties of $f(x)$, in a manner specified by various theorems on the approximate representation of continuous functions by trigonometric sums.* The significance of the relation (9) for the convergence of $T_n(x)$ can be expressed by the general statement

If the sums $t_n(x)$ can be chosen for all values of $n \geq 1$ in such a way that

$$\lim_{n=\infty} n^{2/m} \varepsilon_n = 0,$$

the approximating sum $T_n(x)$ will converge uniformly to the value of $f(x)$, as n becomes infinite, throughout the interval $\alpha \leqq x \leqq \beta$; that is, throughout any interval in which (except possibly for a point set of measure zero) $\varrho(x)$ has a positive lower bound.

If $m = 2$, for example, it is sufficient that $f(x)$ have a continuous derivative† for all values of x; if $m > 2$, it is sufficient that $f(x)$ satisfy a Lipschitz condition, and sufficient, more generally,‡ that $\lim_{\delta=0} \omega(\delta)/\delta^{2/m} = 0$, where $\omega(\delta)$ is the modulus of continuity of $f(x)$, the maximum of $|f(x')-f(x'')|$ for $|x'-x''| \leqq \delta$.

In case the weight-function $\varrho(x)$ is continuous, the above results provide for the convergence of $T_n(x)$—apart from the question of uniform convergence—at all points where $\varrho \neq 0$, if $f(x)$ satisfies suitable conditions.

Although it is perhaps not worth while to use much space here in an attempt to lighten the restrictions on $f(x)$, it may be of some interest to point out one method of reasoning further from the stage that has been reached.§ Let α and β have the same significance as before, and let α_1 and β_1 be two numbers such that $\alpha < \alpha_1 < \beta_1 < \beta$. For $\alpha_1 \leqq x \leqq \beta_1$, the relation (6) may be replaced by the simpler statement that

$$|\tau_n'(x)| \leqq k_1' n\mu,$$

---

* See e. g., D. Jackson, *On approximation by trigonometric sums and polynomials*, these Transactions, vol. 13 (1912), pp. 491-515; *On the approximate representation of an indefinite integral*, etc. (to be cited by the letter D), the same Transactions, vol. 14 (1913), pp. 343-364.

† Cf. the paper D, Theorem IV, p. 351.

‡ Cf. D, p. 350. For the form of the result, cf. the papers A and C.

§ Cf. the paper C, § 6. The method is carried further here than in the passage cited.

where k_1' depends on α, β, α_1, and β_1, but not on anything else. Because of (8), this means that

$$|\tau_n'(x)| \leqq C_1' \varepsilon n^{1+(2/m)}$$

for $n \geqq n_0$, C_1' being independent of x, n, and φ. Let μ_1 be the maximum of $|\tau_n(x)|$ for $\alpha_1 \leqq x \leqq \beta_1$, and let x_1 be a point in (α_1, β_1) such that $|\tau_n(x_1)| = \mu_1$. Let

$$h_1 = \frac{\mu_1}{2 C_1' \varepsilon n^{1+(2/m)}};$$

it is clear that $C_1' \neq 0$. Since $\mu_1 \leqq \mu \leqq C'' \varepsilon n^{2/m}$, it follows that $h_1 \leqq C''/(2C'n)$, and there is a number $n_1 \geqq n_0$, independent of x, n, and φ, such that $h_1 < \frac{1}{2}(\beta_1 - \alpha_1)$ for $n \geqq n_1$. When the last condition is satisfied, at least one of the intervals $(x_1 - h_1, x_1)$, $(x_1, x_1 + h_1)$ will be contained in (α_1, β_1). Throughout one of these intervals, therefore,

$$|\tau_n(x) - \tau_n(x_1)| \leqq \frac{\mu_1}{2}, \qquad |\tau_n(x)| \geqq \frac{\mu_1}{2}.$$

Suppose temporarily that $\varepsilon \leqq \mu_1/4$; as $|\varphi(x)| \leqq \varepsilon$,

$$|\varphi(x) - \tau_n(x)| \geqq \frac{\mu_1}{4}.$$

Since $\varrho(x) \geqq v$ throughout the interval, except possibly for a set of measure zero,

$$\gamma_n \geqq h_1 v \left(\frac{\mu_1}{4}\right)^m = \frac{v \mu_1^{m+1}}{2 \cdot 4^m C_1' \varepsilon n^{1+(2/m)}}.$$

But, as before,

$$\gamma_n \leqq 2\pi V \varepsilon^m.$$

Hence

$$\mu_1 \leqq \sqrt[m+1]{\frac{C_1' \pi V}{v}} \cdot 4 \varepsilon n^{\frac{m+2}{m(m+1)}}.$$

Under the hypothesis that $\varepsilon \leqq \mu_1/4$, therefore, and also, as is recognized *a posteriori*, under the contrary hypothesis,

$$\mu_1 \leqq C_1'' \varepsilon n^{\frac{m+2}{m(m+1)}}$$

for $n \geqq n_1$, the factor C_1'' being independent of x, n, and φ.

This in itself would lead to an improvement in the convergence theorem, as far as the interval (α_1, β_1) is concerned. It is possible, however, to apply the same sort of reasoning again, in an interval (α_2, β_2), where $\alpha_1 < \alpha_2 < \beta_2 < \beta_1$. More generally, let an infinite succession of such intervals be defined, each interior to the preceding; let μ_i be the maximum of $|\tau_n(x)|$ for $\alpha_i \leqq x \leqq \beta_i$, and suppose it has been shown that $\mu_{i-1} \leqq C_{i-1}'' \varepsilon n^{\sigma}$ for $n \geqq n_{i-1}$, where σ is some positive number depending only on m and i, and n_{i-1} is independent of x, n, and φ. By a series of steps analogous to those taken above,

$$|\tau_n'(x)| \leqq C_i' \varepsilon n^{1+\sigma}$$

for $\alpha_i \leqq x \leqq \beta_i$, with C_i' independent of x, n, and φ; for $n \geqq n_i$, a number independent of x, n, and φ, and for $\varepsilon \leqq \mu_i/4$,

$$|\varphi(x) - \tau_n(x)| \geqq \frac{\mu_i}{4}$$

throughout an interval of length

$$h_i = \frac{\mu_i}{2 C_i' \varepsilon n^{1+\sigma}}$$

wholly contained in (α_i, β_i); and so

$$\gamma_n \geqq h_i v \left(\frac{\mu_i}{4}\right)^m = \frac{v \mu_i^{m+1}}{2 \cdot 4^m C_i' \varepsilon n^{1+\sigma}},$$

$$\mu_i \leqq C_i'' \varepsilon n^{\frac{1+\sigma}{m+1}},$$

with C_i'' independent of x, n, and φ, the last relation holding whether the hypothesis $\varepsilon \leqq \mu_i/4$ is verified or not.

Let $\sigma_0 = 2/m$, and $\sigma_i = (1+\sigma_{i-1})/(m+1)$, $i = 1, 2, \cdots$. Then $\mu_i \leqq C_i'' \varepsilon n^{\sigma_i}$ for each value of i, provided that $n > n_i$. It may be verified by induction that

$$\sigma_i = \frac{1}{m} + \frac{1}{m(m+1)^i},$$

the second term on the right approaching zero as i becomes infinite. Let η be an arbitrary positive quantity, and let i be chosen so that $1/[m(m+1)^i] < \eta$.

Let α_0 and β_0 be two arbitrary numbers, subject to the inequalities $\alpha < \alpha_0 < \beta_0 < \beta$. The numbers $\alpha_1, \ldots, \alpha_i, \beta_1, \ldots, \beta_i$ may be chosen so that $\alpha_i = \alpha_0$, $\beta_i = \beta_0$. With the notation $\mu_i = \mu_0$, $C_i'' = C_0''$, $n_i = n_0'$, it is seen that

$$\mu_0 \leqq C_0'' \varepsilon n^{(1/m)+\eta}$$

for $n \geqq n_0'$, where μ_0 now is the maximum of $|\tau_n(x)|$ for $\alpha_0 \leqq x \leqq \beta_0$. By an argument which need not be repeated in detail, it is recognized that $T_n(x)$ converges uniformly to $f(x)$ throughout an arbitrary closed interval interior to (α, β), if there exists a number $\eta > 0$ such that

$$\lim_{\delta=0} \omega(\delta)/\delta^{(1/m)+\eta} = 0,$$

$\omega(\delta)$ being once more the modulus of continuity of $f(x)$. If $f(x)$ satisfies a Lipschitz condition, this requirement will be fulfilled for any value of $m > 1$.

As before, the results are applicable, apart from the question of uniformity, at any point where $\varrho \neq 0$, if ϱ is continuous. The problem of convergence at a point where ϱ vanishes will not be treated at length, but will be touched upon just to the extent of showing that it is not wholly beyond the reach of the methods that have been employed.

Let it be supposed that an interval (α, β) and positive numbers w and s exist so that*

$$\varrho(x) \geqq w(x-\alpha)^s$$

for $\alpha \leqq x \leqq \beta$. Further hypotheses with regard to ϱ are merely that it is everywhere bounded and measurable, and of period 2π. With the same notation as before, and by the same reasoning, there is an interval of length at least $h = 1/(4k'^2 n^2)$, contained in (α, β) for $n \geqq n_0$, throughout which $|\varphi(x) - \tau_n(x)| \geqq \mu/4$, if $\mu \geqq 4\varepsilon$. Let the left-hand end of this interval, x_0 or $x_0 - h$, according to circumstances, be denoted by α'.

* It would come to the same thing if it were assumed that $\rho(x) \geqq w(\beta - x)^s$, or that α is interior to an interval throughout which $\rho(x) \geqq w|x-\alpha|^s$. The essential point is that, for approach from one side at least, ρ has a root of order not higher than s. The most interesting special values of s are presumably $s = 2$, the lowest value consistent with the existence of a derivative for ρ at the point in question, and $s = 1$.

This α' may or may not coincide with α, but in any case $0 \leqq x-\alpha' \leqq x-\alpha$ for $\alpha' \leqq x \leqq \beta$, and so

$$\varrho(x) \geqq w(x-\alpha)^s \geqq w(x-\alpha')^s$$

for $\alpha' \leqq x \leqq \alpha'+h$. Consequently

$$\gamma_n = \int_0^{2\pi} \varrho(x)\,|\varphi(x)-\tau_n(x)|^m\,dx \geqq w\left(\frac{\mu}{4}\right)^m \int_{\alpha'}^{\alpha'+h} (x-\alpha')^s\,dx$$

$$= w\left(\frac{\mu}{4}\right)^m \cdot \frac{h^{s+1}}{s+1} = \frac{w\mu^m}{4^{m+s-1}(s+1)(k'n)^{2s+2}}.$$

On the other hand, as always, $\gamma_n \leqq 2\pi V \varepsilon^m$. So

$$\mu \leqq \bar{C}\varepsilon n^{(2s+2)/m}$$

for $n \geqq n_0$, the values of n_0 and $\bar{C}$ being independent of x, n, and φ, and the conclusion valid even if the hypothesis that $\mu \geqq 4\varepsilon$ is dropped. If $m = 2$ and $s = 1$, $T_n(x)$ will converge to $f(x)$ for $x = \alpha$ (and uniformly for $\alpha \leqq x \leqq \beta$) if $f(x)$ has everywhere a continuous second derivative,* while the existence of a continuous third derivative will be sufficient if $m = s = 2$.

* For the requisite preliminary theorem on the approximate representation of $f(x)$ by means of trigonometric sums, cf. once more the paper D, Theorem IV, p. 351.

THE UNIVERSITY OF MINNESOTA,
MINNEAPOLIS, MINN.